

## Approximation of Lyapunov functions from noisy data

Article (Accepted Version)

Giesl, Peter, Hamzi, Boumediene, Rasmussen, Martin and Webster, Kevin (2020) Approximation of Lyapunov functions from noisy data. *Journal of Computational Dynamics*, 7 (1). pp. 57-81. ISSN 2158-2491

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/87748/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

### **Copyright and reuse:**

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

# APPROXIMATION OF LYAPUNOV FUNCTIONS FROM NOISY DATA

PETER GIESL

Department of Mathematics,  
University of Sussex,  
Falmer, BN1 9QH, UK

BOUMEDIENE HAMZI, MARTIN RASMUSSEN AND KEVIN WEBSTER

Department of Mathematics,  
Imperial College London,  
London SW7 2AZ, UK

(Communicated by the associate editor name)

**ABSTRACT.** Methods have previously been developed for the approximation of Lyapunov functions using radial basis functions. However these methods assume that the evolution equations are known. We consider the problem of approximating a given Lyapunov function using radial basis functions where the evolution equations are not known, but we instead have sampled data which is contaminated with noise. We propose an algorithm in which we first approximate the underlying vector field, and use this approximation to then approximate the Lyapunov function. Our approach combines elements of machine learning/statistical learning theory with the existing theory of Lyapunov function approximation. Error estimates are provided for our algorithm.

Lyapunov function, differential equations, radial basis function, error estimates, reproducing kernel Hilbert space

**1. Introduction.** Ordinary differential equations model large classes of applications such as planetary motion, chemical reactions, population dynamics or consumer behaviour. A breakthrough in the understanding of ordinary differential equations was initiated by Poincaré and Lyapunov in the late 19th century, who developed an approach that embraced the use of topological and geometrical techniques for the study of dynamical systems. A key component of this theory is Lyapunov functions, which can be used to determine the basin of attraction of an asymptotically stable equilibrium.

In general, it is not possible to find an explicit analytical expression for a Lyapunov function associated to a nonlinear differential equation. Many methods have been proposed to numerically construct Lyapunov functions, see [13] for a recent review. These methods include the SOS (sums of squares) method, which constructs a polynomial Lyapunov function by semidefinite optimization [26]. Another method constructs a continuous piecewise affine (CPA) Lyapunov function using linear optimization [16]. A further method is based

---

2010 *Mathematics Subject Classification.* 37B25, 65N15, 37M99.

*Key words and phrases.* Lyapunov function, differential equations, radial basis function, error estimates, reproducing kernel Hilbert space.

B. Hamzi was supported by Marie Curie Fellowships. M. Rasmussen was supported by an EPSRC Career Acceleration Fellowship EP/I004165/1 and K.N. Webster was supported by the EPSRC Grant EP/L00187X/1 and a Marie Skłodowska-Curie Individual Fellowship Grant Number 660616.

on Zubov's equation and computes a solution of this partial differential equation [6]. Lyapunov functions can also be constructed using set oriented methods [15]. The method that is also used in this paper is based on approximating the solution of a PDE using radial basis functions [11]. All these methods to approximate Lyapunov functions rely on the knowledge of the right hand side of the differential equation.

In this paper, we develop a method to approximate Lyapunov functions where the right hand side is unknown, but we have sampled data of the system, which is contaminated by noise. We will first approximate the right hand side of the differential equation, and then use this approximation to approximate the Lyapunov function. Our approach combines and develops previous results from statistical learning theory [29, 30, 31] together with existing methods using radial basis functions [11, 14], which use the framework of reproducing kernel Hilbert spaces (RKHS).

The use of RKHS spaces to approximate important quantities in dynamical systems has previously been exploited by Smale and Zhou to approximate a hyperbolic dynamical system [32]. Bouvrie and Hamzi also use RKHS spaces to approximate some key quantities in control and random dynamical systems [4, 5].

**2. Setting of the Problem and Main Result.** We consider ordinary differential equations of the form

$$\dot{x} = f^*(x), \quad (2.1)$$

where  $f^* : \mathbb{R}^d \rightarrow \mathbb{R}^d$  is a smooth vector field and dot denotes differentiation with respect to time. We define the flow  $\varphi_{f^*} : \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}^d$  by  $\varphi_{f^*}(\psi, t) := x(t)$ , where  $x(t)$  solves (2.1) with  $x(0) = \psi$ .

We assume that (2.1) has a fixed point  $\bar{x}$  that is exponentially asymptotically stable. Define the basin of attraction as  $A(\bar{x}) := \{\psi \in \mathbb{R}^d \mid \lim_{t \rightarrow \infty} \varphi_{f^*}(\psi, t) = \bar{x}\}$ . Note that  $A(\bar{x}) \neq \emptyset$  and  $A(\bar{x})$  is open. Subsets of the basin of attraction can be determined by the use of Lyapunov functions, which are functions decreasing along solutions of (2.1). We consider two types of Lyapunov functions  $V$  and  $T$ , as described in Theorems 3.2 and 3.3 below. These Lyapunov functions satisfy

$$\begin{aligned} \langle \nabla V(x), f^*(x) \rangle_{\mathbb{R}^d} &= -p(x), & x \in A(\bar{x}), \\ \langle \nabla T(x), f^*(x) \rangle_{\mathbb{R}^d} &= -\bar{c}, & x \in A(\bar{x}) \setminus \{\bar{x}\}, \end{aligned}$$

where  $p$  is a smooth function with  $p(x) > 0$  for  $x \neq \bar{x}$  and  $p(\bar{x}) = 0$ , and  $\bar{c}$  is a positive constant. The scalar products on the left hand sides are called the orbital derivatives of  $V$  and  $T$  with respect to (2.1), which are the derivatives of  $V$  and  $T$  along solutions of (2.1). The orbital derivatives of  $V$  and  $T$  are negative, which implies that  $V$  and  $T$  are decreasing along solutions.

We assume that the function  $f^*$  is unknown, but we have sampled data of the form  $(x_i, y_i)$  in  $X \times \mathbb{R}^d$ ,  $i = 1, \dots, m$ , with

$$y_i = f^*(x_i) + \eta_{x_i} \quad (2.2)$$

We assume that the one-dimensional random variables  $\eta_{x_i}^k \in \mathbb{R}^d$ , where  $i = 1, \dots, m$  and  $k = 1, \dots, d$ , are independent random variables drawn from a probability distribution with zero mean and variance  $(\sigma_{x_i}^k)^2$  bounded by  $\sigma^2$ . Here  $X$  is a nonempty and compact subset of  $\mathbb{R}^d$  with  $C^1$  boundary.

In §5 we provide an algorithm to approximately reconstruct the Lyapunov functions  $V$  and  $T$  by functions  $\hat{V}$  and  $\hat{T}$ . The following main theorem provides error estimates in a compact set  $\mathcal{D} \subset A(\bar{x}) \cap X$ , which depend on the density of the data, measured by two key

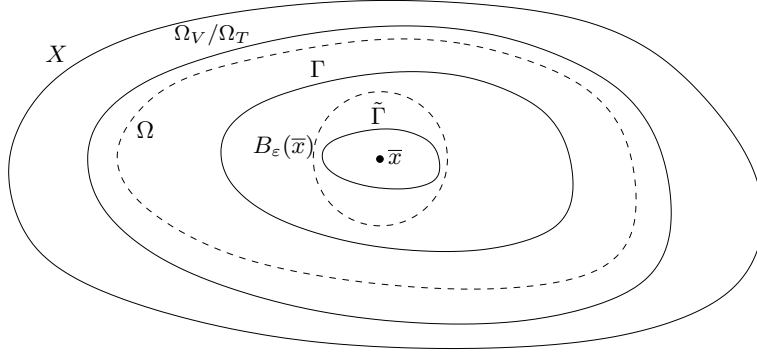


FIGURE 1. Domains and sets used in the statement and proof of Theorem 2.1. The dotted lines show the boundary of the set  $\mathcal{D} = \Omega \setminus B_\varepsilon(\bar{x})$  where the Lyapunov functions are approximated. We also have  $\Omega_V, \Omega_T \subset A(\bar{x})$ .

quantities: the fill distance of the data  $h_{\mathbf{x}}$  (see Definition 6.3) and the norm of the volume weights  $\mathbf{w}$  corresponding to the Voronoi tessellation of the data (see Definition 5.1).

**Theorem 2.1.** Consider (2.1) such that  $f^* \in C^{\nu_1}(\mathbb{R}^d, \mathbb{R}^d)$  with  $\nu_1 \geq (3d+7)/2$  if  $d$  is odd, or  $\nu_1 \geq (3d+12)/2$  if  $d$  is even. Let  $\tau_1, \tau_2 \in \mathbb{R}$  and  $k_1, k_2 \in \mathbb{N}$  be such that  $\tau_1 = k_1 + (d+1)/2$  with  $\lceil \tau_1 \rceil = \nu_1$ , and  $k_2 = k_1 - (d+2)$  (if  $d$  is odd) or  $k_2 = k_1 - (d+3)$  (if  $d$  is even). Define  $\tau_2 := k_2 + (d+1)/2$ .

Let  $\Omega \subset A(\bar{x})$  be a compact set and  $\mathcal{D} := \Omega \setminus B_\varepsilon(\bar{x}) \subset X$ , with  $\varepsilon > 0$  small enough so that  $\mathcal{D} \neq \emptyset$ . For  $h_{\mathbf{x}}$ ,  $\|\mathbf{w}\|_{\mathbb{R}^m}$  and  $h_{\mathbf{q}}$  sufficiently small, the following holds:

1. For every  $0 < \delta < 1$ , the reconstruction  $\hat{V}$  of the Lyapunov function  $V$  defined in Theorem 3.2 satisfies the following estimate with probability  $1 - \delta$ :

$$\left\| \langle \nabla \hat{V}, f^* \rangle_{\mathbb{R}^d} - \langle \nabla V, f^* \rangle_{\mathbb{R}^d} \right\|_{L^\infty(\mathcal{D})} \leq C \|V\|_{W_2^{\tau_2}(\Omega_V)} \left( h_{\mathbf{q}}^{k_2 - \frac{1}{2}} + \frac{\|\mathbf{w}\|_{\mathbb{R}^m}}{\lambda \sqrt{\delta}} + \lambda^{r - \frac{3}{2}} h_{\mathbf{x}} + \lambda^{r - \frac{1}{2}} \right), \quad (2.3)$$

where  $\Omega_V \supset \mathcal{D}$  is a certain compact subset of  $A(\bar{x})$ , and  $\frac{1}{2} < r \leq 1$ .

2. For every  $0 < \delta < 1$ , the reconstruction  $\hat{T}$  of the Lyapunov function  $T$  defined in Theorem 3.3 satisfies the following estimates with probability  $1 - \delta$ :

$$\left\| \langle \nabla \hat{T}, f^* \rangle_{\mathbb{R}^d} - \langle \nabla T, f^* \rangle_{\mathbb{R}^d} \right\|_{L^\infty(\mathcal{D})} \leq C \|T\|_{W_2^{\tau_2}(\Omega_T)} \left( h_{\mathbf{q}}^{k_2 - \frac{1}{2}} + \frac{\|\mathbf{w}\|_{\mathbb{R}^m}}{\lambda \sqrt{\delta}} + \lambda^{r - \frac{3}{2}} h_{\mathbf{x}} + \lambda^{r - \frac{1}{2}} \right), \quad (2.4)$$

$$\left\| \hat{T} - T \right\|_{L^\infty(\Gamma)} \leq C h_{\mathbf{q}}^{k_2 + \frac{1}{2}} \|T\|_{W_2^{\tau_2}(\Omega_T)}, \quad (2.5)$$

where  $\Gamma$  is a non-characteristic hypersurface on which  $T$  has defined values (see Definition 3.4),  $\Omega_T \supset \mathcal{D}$  is a certain compact subset of  $A(\bar{x})$ , and  $\frac{1}{2} < r \leq 1$ .

The main point is that the expressions on the right hand side of (2.3)–(2.5) can be made arbitrarily small as the data density increases and for suitably chosen  $\lambda$  (see equation (6.13)). Therefore the orbital derivative of our Lyapunov function approximations  $\hat{V}$

and  $\hat{T}$  become arbitrarily close in the infinity norm to those of  $V$  and  $T$  respectively. Estimate (2.3) implies that the orbital derivative of  $\hat{V}$  will be negative in  $\mathcal{D}$  (which does not contain a small neighbourhood of the equilibrium  $\bar{x}$ ), since  $\langle \nabla V(x), f^*(x) \rangle_{\mathbb{R}^d} = -p(x)$  where  $p$  is a positive definite function (see Theorem 3.2). The analogous statement is true for  $\hat{T}$ , since  $\langle \nabla T(x), f^*(x) \rangle_{\mathbb{R}^d} = -\bar{c} < 0$ . In principle the neighbourhood  $B_\varepsilon(\bar{x})$  can shrink as the data density increases (as  $h_x$  and  $\|\mathbf{w}\|_{\mathbb{R}^m}$  tend to zero).

The above estimate contains  $\lambda > 0$  as a regularisation parameter of our algorithm, and  $h_q$  as the fill distance of a set of sampled points in  $\Omega_V$  (resp.  $\Omega_T$ ) of our choosing. Similarly,  $h_{\bar{q}}$  is the fill distance of a set of sampled points on  $\Gamma$ , which we are able to choose. The constants in the above estimates depend on  $d, \sigma$ , the choice of function spaces for approximation and the vector field  $f^*$ .

The rest of the paper is organised as follows. In §3 we provide the converse theorems for the Lyapunov functions  $V$  and  $T$ . In §4 we set out the framework for the function spaces that are used to approximate the Lyapunov functions, as well as previous results on the approximation of Lyapunov functions when the right hand side of (2.1) is known. The algorithms themselves that are used to compute  $\hat{V}$  and  $\hat{T}$  are detailed in §5. In §6 we provide an estimate for our approximation of the right hand side of (2.1), which is then used in the proof of Theorem 2.1 in §7.

**3. Converse theorems for Lyapunov functions.** The concept of a Lyapunov function dates back to 1893, where Lyapunov introduced these functions for the stability analysis of an equilibrium for a given differential equation, without the explicit knowledge of the solutions [19]. Many converse theorems have been proved that guarantee the existence of a Lyapunov function under certain conditions, see [11, 18, 21] for an overview. Massera [22] provided the first main converse theorem for  $C^1$  vector fields where  $A(\bar{x}) = \mathbb{R}^d$ , with further developments by several authors to prove the existence of smooth Lyapunov functions under weak smoothness assumptions on the right hand side (see e.g. [17, 20, 34]).

The existence of a Lyapunov function for system (2.1) with given values of the orbital derivative has been shown by Bhatia [2, 3], as stated in the following theorems (see also [11]). We also refer to [12] for a proof that the conditions on the function  $p$  given here are sufficient to define the Lyapunov function  $V$ , in contrast to the conditions given in [11]. First we make the following definition.

**Definition 3.1.** A continuous function  $\alpha : [0, \infty) \rightarrow [0, \infty)$  is a class  $\mathcal{K}$  function if  $\alpha(0) = 0$  and  $\alpha$  is strictly monotonically increasing.

**Theorem 3.2.** Consider the autonomous system of differential equations  $\dot{x} = f^*(x)$ , where  $f^* \in C^{\nu_1}(\mathbb{R}^d, \mathbb{R}^d)$ ,  $\nu_1 \geq 1$ ,  $d \in \mathbb{N}$ . We assume the system to have an exponentially asymptotically stable equilibrium  $\bar{x}$ . Let  $p \in C^{\nu_1}(\mathbb{R}^d, \mathbb{R})$  be a function with the following properties:

1.  $p(x) > 0$  for  $x \neq \bar{x}$ , and  $p(\bar{x}) = 0$ .
2. There is a class  $\mathcal{K}$  function  $\alpha$  such that  $p(x - \bar{x}) \geq \alpha(\|x - \bar{x}\|_2)$  for all  $x \in \mathbb{R}^d$ .

Then there exists a Lyapunov function  $V \in C^{\nu_1}(A(\bar{x}), \mathbb{R})$  (where  $A(\bar{x})$  is the basin of attraction of  $\bar{x}$ ), such that

$$\langle \nabla V(x), f^*(x) \rangle_{\mathbb{R}^d} = -p(x) \quad (3.1)$$

holds for all  $x \in A(\bar{x})$ . The Lyapunov function  $V$  is uniquely defined up to a constant.

We may also choose  $p(x)$  to be a positive constant in equation (3.1) to obtain a Lyapunov function  $T$ , defined on  $A(\bar{x}) \setminus \{\bar{x}\}$ , for which  $\lim_{x \rightarrow \bar{x}} T(x) = -\infty$ .

**Theorem 3.3.** *Consider the autonomous system of differential equations  $\dot{x} = f^*(x)$ , where  $f^* \in C^{\nu_1}(\mathbb{R}^d, \mathbb{R}^d)$ ,  $\nu_1 \geq 1$ ,  $d \in \mathbb{N}$ . We assume the system to have an exponentially asymptotically stable equilibrium  $\bar{x}$ . Then for all  $\bar{c} \in \mathbb{R}^+$ , there exists a Lyapunov function  $T \in C^{\nu_1}(A(\bar{x}) \setminus \{\bar{x}\}, \mathbb{R})$  such that*

$$\langle \nabla T(x), f^*(x) \rangle_{\mathbb{R}^d} = -\bar{c}. \quad (3.2)$$

Moreover,  $\lim_{x \rightarrow \bar{x}} T(x) = -\infty$ .

The Lyapunov function  $T$  will be uniquely defined if its values are prescribed on a non-characteristic hypersurface  $\Gamma \subset A(\bar{x})$  [11] by a function  $\xi_T \in C^{\nu_1}(\Gamma, \mathbb{R})$ ; that is,  $T(x) = \xi_T(x)$  for  $x \in \Gamma$ .

**Definition 3.4** (Non-characteristic hypersurface). Consider  $\dot{x} = f^*(x)$ , where  $f^* \in C^{\nu_1}(\mathbb{R}^d, \mathbb{R}^d)$ ,  $\nu_1 \geq 1$ ,  $d \in \mathbb{N}$ . Let  $h \in C^{\nu_1}(\mathbb{R}^d, \mathbb{R})$  and recall  $\varphi_{f^*} : \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}^d$  denotes the flow mapping. The set  $\Gamma \subset \mathbb{R}^d$  is called a non-characteristic hypersurface if

1.  $\Gamma$  is compact,
2.  $h(x) = 0$  if and only if  $x \in \Gamma$ ,
3.  $h'(x) = \langle h(x), f^*(x) \rangle_{\mathbb{R}^d} < 0$  holds for all  $x \in \Gamma$ , and
4. for each  $x \in A(\bar{x}) \setminus \{\bar{x}\}$  there is a time  $\theta(x) \in \mathbb{R}$  such that  $\varphi_{f^*}(x, \theta(x)) \in \Gamma$ .

An example of a non-characteristic hypersurface is the level set of a (local) Lyapunov function, see [11]. In what follows we assume that we have chosen a non-characteristic hypersurface  $\Gamma$  together with a function  $\xi_T \in C^{\nu_1}(\Gamma, \mathbb{R})$  so that  $T$  is uniquely defined.

**Remark 3.5.** The Lyapunov function  $V$  is smooth and defined on the entire basin of attraction  $A(\bar{x})$ . However, its orbital derivative vanishes at the equilibrium  $\bar{x}$ , and therefore estimates that bound the error of the orbital derivative for numerical approximations of  $V$  cannot guarantee negative orbital derivative arbitrarily close to the equilibrium  $\bar{x}$ . On the other hand, the Lyapunov function  $T$  is not even defined at  $\bar{x}$  and is unbounded near the equilibrium. However, its definition has the advantage that it is not required that we know where the equilibrium is.

In our approach to approximate Lyapunov functions directly from data, we will provide estimates for the approximation of both  $V$  and  $T$ . The strategy is to first approximate  $f^*$  from the data, and use this in turn to approximate the Lyapunov function.

#### 4. Function spaces and approximation theorems for Lyapunov functions.

**4.1. Reproducing kernel Hilbert spaces.** The function spaces that we use to search for our approximations to both  $f^*$  and the Lyapunov functions  $V$  and  $T$  will be reproducing kernel Hilbert spaces (RKHS). For a survey of the main properties of RKHS spaces mentioned in this section, we refer to [7].

In order to define an RKHS function space we first fix a continuous, symmetric, positive definite function (a “kernel”)  $K : X \times X \rightarrow \mathbb{R}$ , and set  $K_x := K(\cdot, x)$ . Define the Hilbert space  $\mathcal{H}_K$  by first considering all finite linear combinations of functions  $K_x$ , that is  $\sum_{x_i \in X} a_i K_{x_i}$  with finitely many  $a_i \in \mathbb{R}$  nonzero. An inner product  $\langle \cdot, \cdot \rangle_K$  on this space is defined by

$$\langle K_{x_i}, K_{x_j} \rangle_K := K(x_i, x_j)$$

and extending linearly. One takes the completion to obtain  $\mathcal{H}_K$ .

Alternatively, an equivalent definition of an RKHS is as a Hilbert space of real-valued functions on  $X$  for which the evaluation functional  $\delta_x(f) := f(x)$  is continuous for all  $x \in X$ .

Finite dimensional subspaces of  $\mathcal{H}_K$  can also be naturally defined by taking a finite number of points  $\mathbf{x} := \{x_1, \dots, x_m\} \subset X$  and considering the linear span

$$\mathcal{H}_{K,\mathbf{x}} := \text{sp}\{K_x : x \in \mathbf{x}\}.$$

In practice we will seek functions in these finite dimensional subspaces as approximations for  $f^*$ .

Within these Hilbert spaces the reproducing property holds:

$$\langle K_x, f \rangle_K = f(x), \quad \forall f \in \mathcal{H}_K. \quad (4.1)$$

If we denote  $\kappa := \sqrt{\sup_{x \in X} K(x, x)}$ , then  $\mathcal{H}_K \subset C(X)$  and it follows that

$$\|f\|_{L^\infty(X)} \leq \kappa \|f\|_K, \quad \forall f \in \mathcal{H}_K. \quad (4.2)$$

The RKHS  $\mathcal{H}_K$  can also be defined by means of an integral operator. Let  $\rho$  be any (finite) strictly positive Borel measure on  $X$  (e.g. Lebesgue measure) and  $L_\rho^2(X)$  be the Hilbert space of square integrable functions on  $X$ . Then define the linear operator  $L_K : L_\rho^2(X) \rightarrow C(X)$  by

$$(L_K f)(x) = \int_X K(x, y) f(y) d\rho(y). \quad (4.3)$$

When composed with the inclusion  $C(X) \hookrightarrow L_\rho^2(X)$  we obtain an operator from  $L_\rho^2(X)$  to  $L_\rho^2(X)$ , which we also denote by  $L_K$ .  $L_K$  is then a self-adjoint compact operator, and it is also positive if the kernel  $K$  is positive definite. Also the map

$$L_K^{1/2} : L_\rho^2(X) \rightarrow \mathcal{H}_K$$

defines an isomorphism of Hilbert spaces.  $L_K^{1/2}$  is well defined as an operator on  $L_\rho^2(X)$  in the sense that  $L_K = L_K^{1/2} \circ L_K^{1/2}$ .

**4.2. Sobolev space RKHS.** In this paper we will work with reproducing kernel Hilbert spaces that are Sobolev spaces. Given the open domain  $\mathcal{B} \subset \mathbb{R}^d$ , for  $k \in \mathbb{N}_0$ ,  $1 \leq p < \infty$ , the Sobolev space  $W_p^k(\mathcal{B})$  consists of all functions  $f$  with weak derivatives  $D^\alpha f \in L^p(\mathcal{B})$ ,  $|\alpha| \leq k$ . We also use the following notation to define the (semi-)norms

$$|f|_{W_p^k(\mathcal{B})} = \left( \sum_{|\alpha|=k} \|D^\alpha f\|_{L^p(\mathcal{B})}^p \right)^{1/p}, \quad \|f\|_{W_p^k(\mathcal{B})} = \left( \sum_{|\alpha| \leq k} \|D^\alpha f\|_{L^p(\mathcal{B})}^p \right)^{1/p}.$$

With  $p = \infty$  the norms are defined in the natural way:

$$|f|_{W_\infty^k(\mathcal{B})} = \sup_{|\alpha|=k} \|D^\alpha f\|_{L^\infty(\mathcal{B})} \quad \text{and} \quad \|f\|_{W_\infty^k(\mathcal{B})} = \sup_{|\alpha| \leq k} \|D^\alpha f\|_{L^\infty(\mathcal{B})}.$$

We will also use fractional order Sobolev spaces. For a detailed discussion see e.g. [1].

The Sobolev embedding theorem states that for  $\tau > d/2$ ,  $W_2^\tau(\mathbb{R}^d)$  can be embedded into  $C(\mathbb{R}^d)$ , and therefore it follows that  $W_2^\tau(\mathbb{R}^d)$  is a RKHS, using the fact that the pointwise evaluation functional is then continuous. Several kernel functions are known to generate RKHS spaces that are norm-equivalent to Sobolev spaces [24, 25]. We will choose to work with the Wendland functions [35]. These are positive definite, compactly supported radial basis function kernels that are represented by a univariate polynomial on their support.

**Definition 4.1** (Wendland function). Let  $l \in \mathbb{N}$ ,  $k \in \mathbb{N}_0$ . We define by recursion

$$\begin{aligned} \psi_{l,0}(r) &= (1-r)_+^l \\ \text{and } \psi_{l,k+1}(r) &= \int_r^1 t \psi_{l,k}(t) dt \end{aligned}$$



for  $r \in \mathbb{R}_0^+$ . Here,  $x_+ = x$  for  $x \geq 0$  and  $x_+ = 0$  for  $x < 0$ .

Setting  $l := \lfloor \frac{d}{2} \rfloor + k + 1$ , the Wendland functions are characterised by a smoothness index  $k \in \mathbb{N}$ , and belong to  $C^{2k}(\mathbb{R}^d)$ . For a domain  $D \subset \mathbb{R}^d$  with a Lipschitz boundary, the Wendland function kernel is given by  $K(x, y) := \psi_{l,k}(c\|x - y\|_{\mathbb{R}^d})$ ,  $c > 0$ , for  $x, y \in D$ . The Wendland function kernel generates an RKHS consisting of the same functions as the Sobolev space  $W_2^\tau(D)$  with  $\tau = k + (d + 1)/2$ , with an equivalent norm [36, Corollary 10.48]. Therefore the generated Sobolev space is of integer order when  $d$  is odd, and integer plus one half when  $d$  is even.

From now on we shall use RKHS spaces generated by Wendland function kernels. We will use two such RKHS spaces for the two parts of our algorithm: to approximate the vector field  $f^*$  in (2.1) we use the space  $\mathcal{H}_{K^1}$  defined on  $X$ , corresponding to the Wendland function kernel  $K^1$  with smoothness index  $k_1$ , such that  $\tau_1 = k_1 + (d + 1)/2$  with  $\lceil \tau_1 \rceil = \nu_1$ . Then  $\mathcal{H}_{K^1}$  is norm-equivalent to  $W_2^{\tau_1}(X)$ . In this case, when  $d$  is odd we have that  $\lceil \tau_1 \rceil = \tau_1$ , and the assumption that  $\nu_1 \geq (3d + 7)/2$  implies that  $k_1 \geq d + 3$ . When  $d$  is even,  $\lceil \tau_1 \rceil = \tau_1 + 1/2$ , and the assumption that  $\nu_1 \geq (3d + 12)/2$  gives  $k_1 \geq d + 5$  (cf. Theorem 2.1).

In the second part of our algorithm we approximate the Lyapunov function. For this, we use the RKHS space  $\mathcal{H}_{K^2}$  defined on  $\Omega_V$  (resp.  $\Omega_T$ ) corresponding to the Wendland function kernel  $K^2$  with smoothness index  $k_2$ , such that  $k_2 = k_1 - (d + 2)$  if  $d$  is odd, or  $k_2 = k_1 - (d + 3)$  if  $d$  is even. Correspondingly, this implies that  $k_2 \geq 1$  when  $d$  is odd, and  $k_2 \geq 2$  when  $d$  is even. Here,  $\mathcal{H}_{K^2}$  is norm-equivalent to  $W_2^{\tau_2}(\Omega_V)$  (resp.  $W_2^{\tau_2}(\Omega_T)$ ), where  $\tau_2 := k_2 + (d + 1)/2$ .

These function spaces consist of smooth functions as a consequence of the following generalised Sobolev inequality (for a proof, see e.g. [9, Chapter 5.7, Theorem 6]).

**Lemma 4.2.** *Let  $\mathcal{B} \subset \mathbb{R}^d$  be a bounded open set with  $C^1$  boundary. For  $u \in W_2^m(\mathcal{B})$  where  $m > d/2$ , we have*

$$\|u\|_{C^{m-\lfloor \frac{d}{2} \rfloor-1, \gamma}(\mathcal{B})} \leq C \|u\|_{W_2^m(\mathcal{B})}, \quad (4.4)$$

where  $\gamma = \frac{1}{2}$  if  $d$  is odd, and  $\gamma$  is any element in  $(0, 1)$  if  $d$  is even.

**Corollary 1.** *For  $f \in \mathcal{H}_{K^1}$  and  $g \in \mathcal{H}_{K^2}$ , we have*

$$\|f\|_{C^{k_1}(X)} \leq C \|f\|_{K^1}, \quad (\text{when } d \text{ is odd}), \quad (4.5)$$

$$\|f\|_{C^{k_1-1}(X)} \leq C \|f\|_{K^1}, \quad (\text{when } d \text{ is even}), \quad (4.6)$$

$$\|g\|_{C^1(\Omega_V/\Omega_T)} \leq C \|g\|_{K^2}. \quad (4.7)$$

*Proof.* Following the arguments given in [23], there exists a bounded extension operator  $E$  that extends  $f \in W_2^{\tau_1}(X)$  to  $Ef \in W_2^{\tau_1}(\mathbb{R}^d)$  such that  $Eu = u$  on  $X$ . Then, from Lemma 4.2 we have

$$\begin{aligned} \|f\|_{C^{\tau_1-\lfloor \frac{d}{2} \rfloor-1, \gamma}(X)} &= \|Ef\|_{C^{\tau_1-\lfloor \frac{d}{2} \rfloor-1, \gamma}(X)} \leq \|Ef\|_{C^{\tau_1-\lfloor \frac{d}{2} \rfloor-1, \gamma}(\mathbb{R}^d)} \\ &\leq \tilde{C} \|Ef\|_{W_2^{\tau_1}(\mathbb{R}^d)} \leq C \|f\|_{W_2^{\tau_1}(X)}. \end{aligned}$$

Then (4.5) and (4.6) follow from using the norm equivalence of  $\mathcal{H}_{K^1}$  to  $W_2^{\tau_1}(X)$ , and that  $\tau_1 = k_1 + (d + 1)/2$ . Inequality (4.7) follows similarly, also using  $k_2 \geq 1$  when  $d$  is odd, and  $k_2 \geq 2$  when  $d$  is even.  $\square$

**4.3. Generalised interpolant and approximation theorems.** In this section we introduce the generalised interpolant that is used to approximate the Lyapunov functions  $V$  and  $T$ . Consider a general interpolation setting where  $\Omega \subset \mathbb{R}^d$  is a bounded domain having a



Lipschitz boundary. Let  $L$  be a linear differential operator and  $\mathbf{q} := \{q_1, \dots, q_M\} \subset \Omega$  be a set of pairwise distinct points which do not contain any singular points of  $L$ . (A point  $q \in \mathbb{R}^d$  is a singular point of  $L$  if  $\delta_q \circ L = 0$ , see also [14].) We define linear functionals

$$\lambda^j(u) := \delta_{q_j} \circ L(u) = Lu(q_j).$$

**Definition 4.3** (Generalized interpolant). Suppose we have values  $Lu(q_i) = \beta_i$  for  $i = 1, \dots, M$  for a function  $u : \Omega \rightarrow \mathbb{R}$ . Given a sufficiently smooth kernel  $K(\cdot, \cdot)$ , define the generalised interpolant as

$$s = \sum_{j=1}^M \alpha_j (\delta_{q_j} \circ L)^y K(\cdot, y),$$

where  $(\delta_{q_j} \circ L)^y$  is the linear function applied to one argument of the kernel. The coefficient vector  $\alpha$  is the solution of  $A\alpha = \beta = (\beta_i)$  with the interpolation matrix  $A \in \mathbb{R}^{M \times M}$  given by

$$(A)_{ij} = (\delta_{q_i} \circ L)^x (\delta_{q_j} \circ L)^y K(x, y).$$

**Remark 4.4.** According to the above definition, we have  $Lu(q_i) = \beta_i$ . In addition, it can be shown that the generalised interpolant above is the unique norm-minimal interpolant in  $\mathcal{H}_K$ , see [11, 14]. The matrix  $A$  is guaranteed to be invertible due to our choice of the Wendland function kernel  $K$  [11, Section 3.2.2], and since  $\mathbf{q}$  does not contain any singular points.

We conclude this section by citing the following theorem from [14], which provides convergence estimates for approximating Lyapunov functions  $V_g$  (resp.  $T_g$ ) with the generalised interpolants  $s_1$  (resp.  $s_2$ ) as above, for a known dynamical system  $\dot{x} = g(x)$ .

**Theorem 4.5.** Let  $\nu_2 := \lceil \tau_2 \rceil$  with  $\tau_2 = k_2 + (d+1)/2$ , where  $k_2$  is the smoothness index of the compactly supported Wendland function. Let  $k_2 > 1/2$  if  $d$  is odd or  $k_2 > 1$  if  $d$  is even. Consider the dynamical system defined by the ordinary differential equation  $\dot{x} = g(x)$ , where  $g \in C^{\nu_2}(\mathbb{R}^d, \mathbb{R}^d)$ . Let  $\bar{x} \in \mathbb{R}^d$  be an equilibrium such that the real parts of all eigenvalues of  $Dg(\bar{x})$  are negative, and suppose  $g$  is bounded in  $A(\bar{x})$ .

Let  $\Gamma$  be a non-characteristic hypersurface as in Definition 3.4, with  $h \in C^{\nu_2}(\mathbb{R}^d, \mathbb{R})$  and  $\xi_T \in C^{\nu_2}(\Gamma, \mathbb{R})$ . Let  $V_g \in W_2^{\tau_2}(A(\bar{x}), \mathbb{R})$ ,  $T_g \in W_2^{\tau_2}(A(\bar{x}) \setminus \{\bar{x}\}, \mathbb{R})$  be the Lyapunov functions of Theorems 3.2 and 3.3 for the system  $\dot{x} = g(x)$ , with  $T_g(x) = \xi_T(x)$  for  $x \in \Gamma$ .

Given pairwise distinct sets of points  $\mathbf{q} := (q_i)_{i=1}^M$  in  $A(\bar{x})$  (not containing  $\bar{x}$ ) and  $\tilde{\mathbf{q}} := (q_i)_{i=M+1}^{M+N}$  in  $\Gamma$ , and let  $\Omega \subset A(\bar{x})$  be a bounded domain with Lipschitz boundary, with  $\mathbf{q} \subset \Omega$ . Let  $s_1$  and  $s_2$  be the generalised interpolants satisfying

$$\begin{aligned} \lambda_g^i(s_1) &:= \langle \nabla s_1(q_i), g(q_i) \rangle_{\mathbb{R}^d} = -p(q_i), & i = 1, \dots, M, \\ \lambda_g^i(s_2) &:= \langle \nabla s_1(q_i), g(q_i) \rangle_{\mathbb{R}^d} = -\bar{c}, & i = 1, \dots, M, \\ \lambda^{i,0}(s_2) &:= s_2(q_i) = \xi_T(q_i), & i = M+1, \dots, M+N, \end{aligned}$$

and let the fill distance of the set of points  $\mathbf{q}$  in  $\Omega$  and  $\tilde{\mathbf{q}}$  in  $\Gamma$  be  $h_{\mathbf{q}}$  and  $h_{\tilde{\mathbf{q}}}$  respectively. Then for  $h_{\mathbf{q}}, h_{\tilde{\mathbf{q}}}$  sufficiently small, the following estimates hold:

$$\|\langle \nabla s_1, g \rangle_{\mathbb{R}^d} - \langle \nabla V_g, g \rangle_{\mathbb{R}^d}\|_{L^\infty(\Omega)} \leq Ch_{\mathbf{q}}^{k-\frac{1}{2}} \|V_g\|_{W_2^{\tau_2}(\Omega)}, \quad (4.8)$$

$$\|\langle \nabla s_2, g \rangle_{\mathbb{R}^d} - \langle \nabla T_g, g \rangle_{\mathbb{R}^d}\|_{L^\infty(\Omega)} \leq Ch_{\mathbf{q}}^{k-\frac{1}{2}} \|T_g\|_{W_2^{\tau_2}(\Omega)}, \quad (4.9)$$

$$\|s_2 - T_g\|_{L^\infty(\Gamma)} \leq Ch_{\tilde{\mathbf{q}}}^{k+\frac{1}{2}} \|T_g\|_{W_2^{\tau_2}(\Omega)}. \quad (4.10)$$

**5. The Algorithm.** Here we present the algorithm for which the estimate given in Theorem 2.1 holds. The algorithm is actually split into two parts. The first part computes  $f_{\mathbf{z},\lambda}$  as an approximation to  $f^*$  (Algorithm 1), and the second part computes  $\hat{V}$  or  $\hat{T}$  as an approximation to the Lyapunov functions  $V$  or  $T$  respectively given in Theorems 3.2 and 3.3 (Algorithm 2A or 2B). As discussed in §4.2, we will use two RKHS spaces  $\mathcal{H}_{K^1}$  and  $\mathcal{H}_{K^2}$  for the two parts of the algorithm, corresponding to the Wendland function kernels  $K^1$  and  $K^2$  with smoothness indices  $k_1$  and  $k_2$  respectively. We recall that the smoothness indices are chosen such that  $\tau_1 = k_1 + (d+1)/2$  with  $\lceil \tau_1 \rceil = \nu_1$ , and  $k_2 = k_1 - (d+3)/2$  if  $d$  is odd, or  $k_2 = k_1 - (d+4)/2$  if  $d$  is even.

Recall that our sampled data values  $\mathbf{z} := (x_i, y_i)_{i=1}^m \in (X \times \mathbb{R}^d)^m$  take the form  $y_i = f^*(x_i) + \eta_{x_i}$ , with  $\eta_x \in \mathbb{R}^d$  a random variable drawn from a probability distribution with zero mean and variance  $\sigma_x^2$ .

Our approximation scheme for  $f^*$  employs a regularised least squares algorithm (see e.g. [10] and its references) to approximate each component  $f^{*,k}$ ,  $k = 1, \dots, d$ . We also introduce a weighting  $\mathbf{w} = \{w_{x_i}\}_{i=1}^m$  corresponding to the Voronoi tessellation associated with the points  $\{x_i\}_{i=1}^m$  [33].

**Definition 5.1** (Voronoi tessellation). Let  $X \subset \mathbb{R}^d$  be compact. For a set of pairwise distinct points  $\mathbf{x} := \{x_i\}_{i=1}^m \in X^m$ , the Voronoi tessellation is the collection of pairwise disjoint open sets  $\mathcal{V}_i(\mathbf{x})$ ,  $i = 1, \dots, m$  defined by

$$\mathcal{V}_i(\mathbf{x}) = \{y \in X \mid \|x_i - y\|_{\mathbb{R}^d} < \|x_j - y\|_{\mathbb{R}^d}, \text{ if } i \neq j\}.$$

The weighting  $\mathbf{w} = \{w_{x_i}\}_{i=1}^m$  is then defined by  $w_{x_i} = \rho(\mathcal{V}_i(\mathbf{x}))$ , where  $\rho$  is the strictly positive Borel measure from §4.1.

**Algorithm 1.** Fix a regularisation parameter  $\lambda > 0$ , and define  $D_w \in \mathbb{R}^{m \times m}$  as the diagonal matrix with diagonal elements  $w_{x_i}$ ,  $i = 1, \dots, m$ . The approximation  $f_{\mathbf{z},\lambda}$  for  $f^*$  is constructed component-wise. That is, for each  $k = 1, \dots, d$ , we approximate the  $k$ -th component  $f^{*,k}$  by  $f_{\mathbf{z},\lambda}^k \in \mathcal{H}_{K^1}$ , defined by  $f_{\mathbf{z},\lambda}^k = \sum_{i=1}^m a_i K_{x_i}^1$ , where the coefficients  $a := \{a_i\}_{i=1}^m$  may be calculated as the solution to the matrix equation

$$(A_{\mathbf{x}} D_w A_{\mathbf{x}} + \lambda A_{\mathbf{x}})a = A_{\mathbf{x}} D_w y^k.$$

Here  $y^k = (y_i^k)_{i=1}^m$  where  $y_i^k$  is simply the  $k$ -th component of  $y_i \in \mathbb{R}^d$ , and  $A_{\mathbf{x}} \in \mathbb{R}^{m \times m}$  is a symmetric matrix defined by  $(A_{\mathbf{x}})_{i,j} = K^1(x_i, x_j)$ .

We note here that due to our choice of RKHS the matrix  $A_{\mathbf{x}}$  is positive definite [14, Proposition 3.3], and therefore the matrix  $(A_{\mathbf{x}} D_w A_{\mathbf{x}} + \lambda A_{\mathbf{x}})$  is invertible, as it is the sum of two positive definite matrices. The error in our approximation of  $f^*$  by  $f_{\mathbf{z},\lambda}$  is studied in §6. We will show that this error may be bounded in the supremum norm on the domain  $X$ , depending on the density of the data in  $X$  and the choice of regularisation parameter  $\lambda$ .

Once we have our approximation  $f_{\mathbf{z},\lambda}$ , then we construct our Lyapunov function approximation with the generalised interpolant as in Theorem 4.5, where we set  $g = f_{\mathbf{z},\lambda}$ .

Therefore we have the following Algorithm 2A for  $\hat{V}$ , or Algorithm 2B for  $\hat{T}$ . These algorithms involve sampling our approximation  $f_{\mathbf{z},\lambda}$  at a discrete set of points.

**Algorithm 2A** (Approximation of  $V$ ). First run Algorithm 1 on the sampled data set  $(x_i, y_i)_{i=1}^m$  to compute  $f_{\mathbf{z},\lambda} \in (\mathcal{H}_{K^1})^d$ .

Define a set of pairwise distinct points  $\mathbf{q} := (q_i)_{i=1}^M$ , with  $f_{\mathbf{z},\lambda}(q_i) \neq 0$  for all  $i = 1, \dots, M$ . The approximation  $\hat{V} \in \mathcal{H}_{K^2}$  for the Lyapunov function  $V$  (see Theorem 3.2) is given by  $\hat{V} = \sum_{i=1}^M b_i \lambda_{f_{\mathbf{z},\lambda}}^{i,y} K^2(\cdot, y)$ , where  $\lambda_{f_{\mathbf{z},\lambda}}^{i,y}$  is the linear functional  $\lambda_{f_{\mathbf{z},\lambda}}^i$  applied

to one argument of the kernel. The coefficients  $b := \{b_i\}_{i=1}^M$  are given by

$$B_{\mathbf{q}} b = -p.$$

Here  $B_{\mathbf{q}} \in \mathbb{R}^{M \times M}$  is a symmetric matrix defined by  $(B_{\mathbf{q}})_{i,j} = \lambda_{f_{\mathbf{z},\lambda}}^{i,x} \lambda_{f_{\mathbf{z},\lambda}}^{j,y} K^2(x, y)$  and  $p = (p(q_i))_{i=1}^M$ .

As before, the choice of RKHS guarantees that the matrix  $B_{\mathbf{q}}$  will be positive definite, provided that  $f_{\mathbf{z},\lambda}(q_i) \neq 0$  for all  $i = 1, \dots, M$ .

To approximate  $T$ , we assume that a non-characteristic hypersurface for  $f^*$  has been defined as  $\Gamma = \{x \in A(\bar{x}) \setminus \{\bar{x}\} \mid h(x) = 0\}$  according to Definition 3.4, for which  $T(x) = \xi_T(x)$  on  $\Gamma$ ,  $\xi_T \in C^{\nu_1}(\Gamma, \mathbb{R})$ .

**Algorithm 2B** (Approximation of  $T$ ). First run Algorithm 1 on the sampled data set  $(x_i, y_i)_{i=1}^m$  to compute  $f_{\mathbf{z},\lambda} \in (\mathcal{H}_{K^1})^d$ .

Define a set of pairwise distinct points  $\mathbf{q} := (q_i)_{i=1}^{M+N}$ , with  $f_{\mathbf{z},\lambda}(q_i) \neq 0$  for all  $i = 1, \dots, M$ , and  $q_i \in \Gamma$  for  $i = M+1, \dots, M+N$ . The approximation  $\hat{T} \in \mathcal{H}_{K^2}$  for the Lyapunov function  $T$  (see Theorem 3.3) is given by  $\hat{T} = \sum_{i=1}^M c_i \lambda_{f_{\mathbf{z},\lambda}}^{i,y} K^2(\cdot, y) + \sum_{i=M+1}^{M+N} c_i K^2(\cdot, q_i)$ , where the coefficients  $c := \{c_i\}_{i=1}^{M+N}$  are computed as the solution to the matrix equation

$$C_{\mathbf{q},\hat{\mathbf{q}}} c = \beta, \quad \text{where } C_{\mathbf{q},\hat{\mathbf{q}}} := \begin{pmatrix} C & D \\ D^T & C^0 \end{pmatrix} \in \mathbb{R}^{(M+N) \times (M+N)},$$

and the submatrices  $C \in \mathbb{R}^{M \times M}$ ,  $D \in \mathbb{R}^{M \times N}$  and  $C^0 \in \mathbb{R}^{N \times N}$  have elements  $(C)_{i,j} = \lambda_{f_{\mathbf{z},\lambda}}^{i,x} \lambda_{f_{\mathbf{z},\lambda}}^{j,y} K^2(x, y)$ ,  $(D)_{i,j-N} = \lambda_{f_{\mathbf{z},\lambda}}^{i,y} K^2(q_j, y)$  and  $(C^0)_{i-N,j-N} = K^2(q_i, q_j)$ . The vector  $\beta$  is given by  $\beta_i = -\bar{c}$ ,  $i = 1, \dots, M$  and  $\beta_i = \xi_T(q_i)$ ,  $i = M+1, \dots, M+N$ .

It may again be shown that the matrix  $C_{\mathbf{q},\hat{\mathbf{q}}}$  will be positive definite, providing that  $f_{\mathbf{z},\lambda}(q_i) \neq 0$  for all  $i = 1, \dots, M$ , for details see [11, Section 3.2.2].

Our error in the approximations  $\hat{V}$  and  $\hat{T}$  will depend primarily on the error induced by Algorithm 1, which in turn depends on the density of the data, as well as the regularisation parameter. In addition, there will be an error due to the discrete sampling of  $f_{\mathbf{z},\lambda}$  in Algorithms 2A and 2B. This error will depend on the density of the sample points  $\mathbf{q}$  (chosen by the user), which in principle can be entirely independent of the original set of points  $\mathbf{x}$  provided by the data. The overall error is the subject of §7, which will prove the estimate given in Theorem 2.1.

**6. Error estimate for  $\|f_{\mathbf{z},\lambda}^k - f^{*,k}\|_{L^\infty(X)}$ .** In this section we estimate the error  $\|f_{\mathbf{z},\lambda}^k - f^{*,k}\|_{L^\infty(X)}$  for each  $k = 1, \dots, d$ .

We have sampled data of the form  $(x_i, y_i)$  in  $X \times \mathbb{R}^d$ ,  $i = 1, \dots, m$ , with  $y_i = f^*(x_i) + \eta_{x_i}$ . We assume that the one-dimensional random variables  $\eta_{x_i}^k \in \mathbb{R}^d$ , where  $i = 1, \dots, m$  and  $k = 1, \dots, d$ , are independent random variables drawn from a probability distribution with zero mean and variance  $(\sigma_{x_i}^k)^2$  bounded by  $\sigma^2$ .

In order to ease notation, and since each  $f_{\mathbf{z},\lambda}^k$  is calculated independently for each  $k$ , we shall henceforth drop the superscript  $k$  and consider the data to be of the form  $\mathbf{z} := (x_i, y_i)_{i=1}^m \in (X \times \mathbb{R})^m$ . Note that in this section we shall only be working with the RKHS  $\mathcal{H}_{K^1}$ .

The following operator definition will enable convenient function representations (c.f. [29, 30]).

**Definition 6.1** (Sampling Operator). Given a set  $\mathbf{x} := (x_i)_{i=1}^m$  of pairwise distinct points in  $\mathbb{R}^d$ , the sampling operator  $S_{\mathbf{x}} : \mathcal{H}_{K^1} \rightarrow \mathbb{R}^m$  is defined as

$$S_{\mathbf{x}}(f) = (f(x_i))_{i=1}^m.$$

The adjoint operator  $S_{\mathbf{x}}^*$  can also be derived as follows. Let  $c \in \mathbb{R}^m$ , then we have

$$\langle f, S_{\mathbf{x}}^* c \rangle_{K^1} = \langle S_{\mathbf{x}} f, c \rangle_{\mathbb{R}^m} = \sum_{i=1}^m c_i f(x_i) = \left\langle f, \sum_{i=1}^m c_i K_{x_i}^1 \right\rangle_{K^1}, \quad \forall f \in \mathcal{H}_{K^1}.$$

The final equality follows from the reproducing property (4.1). So then  $S_{\mathbf{x}}^* c = \sum_{i=1}^m c_i K_{x_i}^1$  for all  $c \in \mathbb{R}^m$ .

The following Lemma shows that the function  $f_{\mathbf{z}, \lambda}$  calculated in Algorithm 1 is the minimiser of a regularised cost function. We omit the proof, which is similar to that contained in [30, Theorem 1], except for the introduction of the weights  $w$ .

**Lemma 6.2.** *Let*

$$f_{\mathbf{z}, \lambda} := \arg \min_{f \in \mathcal{H}_{K^1}} \left\{ \sum_{i=1}^m w_{x_i} (f(x_i) - y_i)^2 + \lambda \|f\|_{K^1}^2 \right\}. \quad (6.1)$$

*Let  $D_w \in \mathbb{R}^{m \times m}$  be the diagonal matrix with entries  $\{w_{x_i}\}_{i=1}^m$ , and let  $\mathbf{y} := \{y_i\}_{i=1}^m$ . If  $S_{\mathbf{x}}^* D_w S_{\mathbf{x}} + \lambda I$  is invertible, then  $f_{\mathbf{z}, \lambda}$  exists and is given by*

$$f_{\mathbf{z}, \lambda} = J \mathbf{y}, \quad J := (S_{\mathbf{x}}^* D_w S_{\mathbf{x}} + \lambda I)^{-1} S_{\mathbf{x}}^* D_w. \quad (6.2)$$

*Furthermore, if  $f_{\mathbf{z}, \lambda} = \sum_{i=1}^m a_i K_{x_i}^1$ , then the coefficients  $a := \{a_i\}_{i=1}^m$  may be calculated as*

$$a = (A_{\mathbf{x}} D_w A_{\mathbf{x}} + \lambda A_{\mathbf{x}})^{-1} A_{\mathbf{x}} D_w \mathbf{y}, \quad (6.3)$$

*where  $A_{\mathbf{x}} \in \mathbb{R}^{m \times m}$  is the symmetric matrix defined by  $(A_{\mathbf{x}})_{i,j} = K^1(x_i, x_j)$ .*

Our strategy to prove convergence of the estimate  $f_{\mathbf{z}, \lambda}$  to  $f^*$  combines and adapts results contained in [29, 30, 31]. The main difference in our case to the standard assumptions in learning theory is that the data  $\mathbf{z} = (x_i, y_i)_{i=1}^m$  is not necessarily generated from an underlying probability distribution. Instead, our data is generated by the underlying dynamical system (2.1), and the data sites may be situated arbitrarily. They could potentially be chosen deterministically, or indeed may be generated randomly. The assumptions made in [30] correspond to this setting, but we would like to provide estimates in terms of the density of the data. This is why we have introduced the weights  $(w_{x_i})_{i=1}^m$  corresponding to the  $\rho$ -volume Voronoi tessellation (c.f. also [29], where a weighting scheme is introduced in a different setting).

**Definition 6.3** (Fill distance). Let  $\mathcal{C} \subset \mathbb{R}^d$  be a compact set and  $\mathbf{x} := \{x_1, \dots, x_m\}$  be a grid, where  $\mathbf{x} \subset \mathcal{C}$ . We denote the *fill distance* of  $\mathbf{x}$  in  $\mathcal{C}$  as

$$h_{\mathbf{x}} = \max_{y \in \mathcal{C}} \min_{x_i \in \mathbf{x}} \|x_i - y\|.$$

In particular, for all  $y \in \mathcal{C}$  there is a grid point  $x_i \in \mathbf{x}$  such that  $\|y - x_i\| \leq h_{\mathbf{x}}$ .

In order to provide an estimate for  $\|f_{\mathbf{z}, \lambda} - f^*\|_{K^1}$ , we first define  $f_{\mathbf{x}, \lambda}, f_{\lambda} \in \mathcal{H}_{K^1}$  by

$$f_{\mathbf{x}, \lambda} := J(S_{\mathbf{x}} f^*) \quad (6.4)$$

$$\text{and } f_{\lambda} := \arg \min_{f \in \mathcal{H}_{K^1}} \{ \|f - f^*\|_{\rho}^2 + \lambda \|f\|_{K^1}^2 \} \quad (6.5)$$

where  $\|f\|_{\rho}^2 = \int_X |f(x)|^2 d\rho(x)$ . Recall from Section 4.1 that  $\rho$  is a finite strictly positive Borel measure on  $X$ .

The function  $f_{\mathbf{x},\lambda}$  may be seen as a ‘noise-free’ version of  $f_{\mathbf{z},\lambda}$ , such that in the case of no noise, i.e.  $\eta_x = 0$  for all  $x \in X$ , then we would have  $f_{\mathbf{z},\lambda} = f_{\mathbf{x},\lambda}$ . Correspondingly, the function  $f_\lambda$  can be seen as a ‘data-free’ limit of  $f_{\mathbf{x},\lambda}$ , or the limiting function as the data sites  $\mathbf{x}$  become arbitrarily dense in  $X$ . Finally, if  $f^* \in \mathcal{H}_{K^1}$ , then  $f^*$  is the ‘regularisation-free’ version of  $f_\lambda$  – the limit of  $f_\lambda$  as  $\lambda \rightarrow 0$ .

The strategy is to break down the estimate of  $\|f_{\mathbf{z},\lambda} - f^*\|_{L^\infty(X)}$  according to

$$\begin{aligned} \|f_{\mathbf{z},\lambda} - f^*\|_{L^\infty(X)} &= \|f_{\mathbf{z},\lambda} - f_{\mathbf{x},\lambda} + f_{\mathbf{x},\lambda} - f_\lambda + f_\lambda - f^*\|_{L^\infty(X)} \\ &\leq \|f_{\mathbf{z},\lambda} - f_{\mathbf{x},\lambda}\|_{L^\infty(X)} + \|f_{\mathbf{x},\lambda} - f_\lambda\|_{L^\infty(X)} + \|f_\lambda - f^*\|_{L^\infty(X)} \end{aligned}$$

and estimate each of the three terms in the inequality. These three terms correspond to errors incurred by the noise (sample error), the finite set of data sites (integration error) and the regularisation parameter  $\lambda$  (regularisation error).

**6.1. Sample error.** An estimate for  $\|f_{\mathbf{z},\lambda} - f_{\mathbf{x},\lambda}\|_{K^1}^2$  for an unweighted approximation scheme is given in Theorem 2 of [30]. A similar result is given in the following lemma, that incorporates the weights of our scheme. The proof follows that of [30], and here we will just sketch the main adaptations. Importantly, our estimate provides convergence of  $f_{\mathbf{z},\lambda}$  to  $f_{\mathbf{x},\lambda}$  as the data sites  $\mathbf{x}$  become more dense, or as the quantity  $\|\mathbf{w}\|_{\mathbb{R}^m}$  tends to zero.

**Lemma 6.4.** *For  $\lambda > 0$ , for every  $0 < \delta < 1$ , with probability  $1 - \delta$ , we have*

$$\|f_{\mathbf{z},\lambda} - f_{\mathbf{x},\lambda}\|_{L^\infty(X)} \leq \kappa \|f_{\mathbf{z},\lambda} - f_{\mathbf{x},\lambda}\|_{K^1} \leq \frac{\|\mathbf{w}\|_{\mathbb{R}^m} \sigma \kappa^2}{\lambda \sqrt{\delta}} \quad (6.6)$$

where  $\sigma^2 := \sup_{x \in X} \sigma_x^2$  and  $\kappa^2 := \sup_{x \in X} K^1(x, x)$ .

*Proof.* First note that if  $\lambda > 0$  then  $(S_{\mathbf{x}}^* D_w S_{\mathbf{x}} + \lambda I)$  is invertible. This follows since it is the sum of a positive and a strictly positive operator on  $\mathcal{H}_{K^1}$ . Similar to [30, Theorem 2], we have

$$\|S_{\mathbf{x}}^* D_w (y - S_{\mathbf{x}} f^*)\|_{K^1}^2 = \sum_{i,j=1}^m w_{x_i} w_{x_j} \eta_{x_i} \eta_{x_j} K^1(x_i, x_j).$$

Since we have assumed that the  $\eta_x$  random variables are independent with zero mean, we have that

$$\mathbb{E}(\|S_{\mathbf{x}}^* D_w (y - S_{\mathbf{x}} f^*)\|_{K^1}^2) = \sum_{i=1}^m w_{x_i}^2 \sigma_{x_i}^2 K^1(x_i, x_i) \leq \sigma^2 \kappa^2 \sum_{i=1}^m w_{x_i}^2 = \sigma^2 \kappa^2 \|\mathbf{w}\|^2,$$

where the inequality follows from our assumption that  $\sigma^2 := \sup_{x \in X} \sigma_x^2 < \infty$ . Then it follows that

$$\mathbb{E}(\|f_{\mathbf{z},\lambda} - f_{\mathbf{x},\lambda}\|_{K^1}^2) \leq \|(S_{\mathbf{x}}^* D_w S_{\mathbf{x}} + \lambda I)^{-1}\|^2 \|\mathbf{w}\|^2 \sigma^2 \kappa^2.$$

The operator  $(S_{\mathbf{x}}^* D_w S_{\mathbf{x}} + \lambda I)^{-1}$  is estimated analogously to [30, Proposition 1] to obtain

$$\|(S_{\mathbf{x}}^* D_w S_{\mathbf{x}} + \lambda I)^{-1}\| \leq \frac{1}{\lambda}. \quad (6.7)$$

Then we have

$$\mathbb{E}(\|f_{\mathbf{z},\lambda} - f_{\mathbf{x},\lambda}\|_{K^1}^2) \leq \frac{\|\mathbf{w}\|^2 \sigma^2 \kappa^2}{\lambda^2}.$$

Finally, for  $0 < \delta < 1$ , application of the Markov inequality to the random variable  $\|f_{\mathbf{z},\lambda} - f_{\mathbf{x},\lambda}\|_{K^1}^2$  gives

$$\mathbb{P}\left(\|f_{\mathbf{z},\lambda} - f_{\mathbf{x},\lambda}\|_{K^1}^2 \geq \frac{\|\mathbf{w}\|^2 \sigma^2 \kappa^2}{\lambda^2 \delta}\right) \leq \delta.$$

Combining the above together with (4.2) proves the lemma.  $\square$

**6.2. Integration error.** To establish our estimate for the integration error we need to make additional assumptions on the choice of Borel measure  $\rho$ . Namely, we will require that it is strongly continuous (c.f. [27]):

**Definition 6.5.** A Borel measure  $\rho$  is strongly continuous if for all hyperplanes  $H \subset \mathbb{R}^d$ , we have  $\rho(H) = 0$ .

Note that this requirement implies that the boundaries of the Voronoi tessellation have  $\rho$ -measure zero. Lebesgue measure is still an example measure that satisfies all of our assumptions, recall Section 4.1. An estimate for the integration error  $\|f_{\mathbf{x},\lambda} - f_\lambda\|_{L^\infty(X)}$  is given in the following lemma.

**Lemma 6.6.** Let  $\rho$  be a (finite) strictly positive, strongly continuous Borel measure on  $X$ . For  $\lambda > 0$ , we have

$$\|f_{\mathbf{x},\lambda} - f_\lambda\|_{L^\infty(X)} \leq \frac{C}{\lambda} \kappa \|f^* - f_\lambda\|_{K^1} h_{\mathbf{x}} \rho(X). \quad (6.8)$$

*Proof.* It is shown in [8] that the solution to (6.5) for  $\lambda > 0$  is given by

$$f_\lambda = (L_{K^1} + \lambda I)^{-1} L_{K^1} f^*, \quad (6.9)$$

where  $L_{K^1}$  was defined in equation (4.3).

Now we have

$$\begin{aligned} f_{\mathbf{x},\lambda} - f_\lambda &= (S_{\mathbf{x}}^* D_w S_{\mathbf{x}} + \lambda I)^{-1} S_{\mathbf{x}}^* D_w S_{\mathbf{x}} f^* - f_\lambda \\ &= (S_{\mathbf{x}}^* D_w S_{\mathbf{x}} + \lambda I)^{-1} \{S_{\mathbf{x}}^* D_w S_{\mathbf{x}} (f^* - f_\lambda) - (L_{K^1} f^* - L_{K^1} f_\lambda)\} \\ &= (S_{\mathbf{x}}^* D_w S_{\mathbf{x}} + \lambda I)^{-1} \left\{ \sum_{i=1}^m w_{x_i} (f^*(x_i) - f_\lambda(x_i)) K_{x_i}^1 - L_{K^1} (f^* - f_\lambda) \right\} \end{aligned}$$

where the second equality follows from (6.9) and the final equality follows from the definition of  $S_{\mathbf{x}}$  and its adjoint.

Recall that we have chosen the weighting  $\mathbf{w}$  to be equal to the  $\rho$ -volume of the Voronoi tessellation associated to the data sites  $\mathbf{x}$  – that is,  $w_{x_i} = \rho(V_i(\mathbf{x}))$ . Also, since  $\rho$  is strongly continuous it holds that

$$L_{K^1} (f^* - f_\lambda) = \sum_{i=1}^m \int_{V_i(\mathbf{x})} K^1(x, y) (f^*(y) - f_\lambda(y)) d\rho(y)$$

and so

$$\begin{aligned} &\left\| \sum_{i=1}^m w_{x_i} (f^*(x_i) - f_\lambda(x_i)) K_{x_i}^1 - L_{K^1} (f^* - f_\lambda) \right\|_{L^\infty(X)} \\ &= \left\| \sum_{i=1}^m \left\{ w_{x_i} (f^*(x_i) - f_\lambda(x_i)) K_{x_i}^1 - \int_{V_i(\mathbf{x})} K_y^1 (f^*(y) - f_\lambda(y)) d\rho(y) \right\} \right\|_{L^\infty(X)} \\ &\leq \sum_{i=1}^m \int_{V_i(\mathbf{x})} \| (f^*(x_i) - f_\lambda(x_i)) K_{x_i}^1 - (f^*(y) - f_\lambda(y)) K_y^1 \|_{L^\infty(X)} d\rho(y) \\ &\leq C \kappa \|f^* - f_\lambda\|_{K^1} \sum_{i=1}^m \int_{V_i(\mathbf{x})} |x_i - y| d\rho(y) \leq C \kappa \|f^* - f_\lambda\|_{K^1} h_{\mathbf{x}} \rho(X) \end{aligned}$$

The second equality follows from the fact that  $(f^* - f_\lambda)$  and  $K_x$  belong to  $\mathcal{H}_{K^1}$ , and are therefore bounded and Lipschitz on  $X$  (cf. Corollary 1). This, together with (6.7) proves the lemma.  $\square$

**6.3. Regularisation error.** For the regularisation error we recall the following result from [31, Lemma 3] (c.f. also [30, Theorem 4]):

**Lemma 6.7.** *Suppose that  $L_{K^1}^{-r} f^* \in L_\rho^2(X)$  for some  $\frac{1}{2} < r \leq 1$ . Then we have*

$$\|f_\lambda - f^*\|_{K^1} \leq \lambda^{r-\frac{1}{2}} \|L_{K^1}^{-r} f^*\|_{L_\rho^2(X)}, \quad \frac{1}{2} < r \leq 1. \quad (6.10)$$

**6.4. Estimate for  $\|f_{\mathbf{z},\lambda} - f^*\|_{L^\infty(X)}$ .** Altogether, from lemmas 6.4, 6.6 and 6.7 and equation (4.2) we have with probability  $1 - \delta$

$$\|f_{\mathbf{z},\lambda} - f^*\|_{L^\infty(X)} \leq \frac{\|\mathbf{w}\|_{\mathbb{R}^m} \sigma \kappa^2}{\lambda \sqrt{\delta}} + \frac{C \kappa \|f^* - f_\lambda\|_{K^1} h_{\mathbf{x}} \rho(X)}{\lambda} + \kappa \lambda^{r-\frac{1}{2}} \|L_{K^1}^{-r} f^*\|_{L_\rho^2(X)} \quad (6.11)$$

Applying equation (6.10) again yields

$$\|f_{\mathbf{z},\lambda} - f^*\|_{L^\infty(X)} \leq C \left( \frac{\|\mathbf{w}\|_{\mathbb{R}^m}}{\lambda \sqrt{\delta}} + \lambda^{r-\frac{3}{2}} h_{\mathbf{x}} + \lambda^{r-\frac{1}{2}} \right), \quad (6.12)$$

where the constant  $C$  depends on  $f^*$ ,  $d$ ,  $\sigma$  and the choice of RKHS  $\mathcal{H}_{K^1}$ . Now it is clear that the above bound can be made arbitrarily small as  $\|\mathbf{w}\|_{\mathbb{R}^m}$  and  $h_{\mathbf{x}}$  tend to zero, if  $\lambda$  also tends to zero at an appropriate rate. With the choice of regularisation parameter

$$\lambda = \left( \max \left\{ \|\mathbf{w}\|_{\mathbb{R}^m}, h_{\mathbf{x}}^{\frac{2}{3-2r}} \right\} \right)^{\frac{2}{2r+1}}, \quad (6.13)$$

with probability  $1 - \delta$ , we obtain the estimate

$$\|f_{\mathbf{z},\lambda} - f^*\|_{L^\infty(X)} \leq C \left( \max \left\{ \|\mathbf{w}\|_{\mathbb{R}^m} / \sqrt{\delta}, h_{\mathbf{x}} \right\} \right)^{\frac{2r-1}{2r+1}}. \quad (6.14)$$

**7. Proof of Theorem 2.1.** Let  $V \in C^{\nu_1}(A(\bar{x}), \mathbb{R})$  and  $T \in C^{\nu_1}(A(\bar{x}) \setminus \{\bar{x}\}, \mathbb{R})$  be the Lyapunov functions for  $f^*$  as defined in Theorems 3.2 and 3.3. Then we have

$$L_{f^*} V(x) := \langle \nabla V(x), f^*(x) \rangle_{\mathbb{R}^d} = -p(x), \quad \text{for all } x \in A(\bar{x}), \quad (7.1)$$

with  $p(x)$  also defined in Theorem 3.2. Similarly,

$$L_{f^*} T(x) = -\bar{c} \quad \text{for all } x \in A(\bar{x}) \setminus \{\bar{x}\}, \quad (7.2)$$

$$T(x) = \xi_T(x), \quad x \in \Gamma. \quad (7.3)$$

for  $c > 0$ , where  $\Gamma = \{x \in A(\bar{x}) \setminus \{\bar{x}\} \mid h(x) = 0\}$  is a non-characteristic hypersurface according to Definition 3.4 (with  $h \in C^{\nu_1}(\mathbb{R}^d, \mathbb{R})$ ), and  $\xi_T \in C^{\nu_1}(\Gamma, \mathbb{R})$ .

As stated in Theorem 3.2, the Lyapunov function  $V$  is uniquely defined up to a constant. We will fix  $V$  by setting  $V(\bar{x}) = 0$ . The Lyapunov function  $T$  is uniquely defined according to the above properties.

The following Lemma provides an alternative characterisation of the Lyapunov function  $V$ , which will be useful later in the section.

**Lemma 7.1.** *Let  $V \in C^{\nu_1}(A(\bar{x}), \mathbb{R})$  be the uniquely defined Lyapunov function as above, and  $\Gamma = \{x \in A(\bar{x}) \setminus \{\bar{x}\} \mid h(x) = 0\}$  ( $h \in C^{\nu_1}(\mathbb{R}^d, \mathbb{R})$ ) is a non-characteristic hypersurface according to Definition 3.4. Define  $\xi_V \in C^{\nu_1}(\Gamma, \mathbb{R})$  by  $\xi_V(x) := V(x)$  for  $x \in \Gamma$ . Also recall  $\varphi_{f^*}(t, \cdot)$  denotes the flow operator of (2.1), and define the function  $\theta_{f^*} \in C^{\nu_1}(A(\bar{x}) \setminus \{\bar{x}\}, \mathbb{R})$  by  $\varphi_{f^*}(t, x) \in \Gamma \Leftrightarrow t = \theta_{f^*}(x)$ . Then*

$$V(x) = \xi_V(\varphi_{f^*}(\theta_{f^*}(x), x)) + \int_0^{\theta_{f^*}(x)} p(\varphi_{f^*}(\tau, x)) d\tau, \quad x \in A(\bar{x}) \setminus \{\bar{x}\}.$$



*Proof.* It is shown in [11, Theorem 2.38] that the function  $\theta_{f^*}$  is well-defined and is in  $C^{\nu_1}(A(\bar{x}) \setminus \{\bar{x}\}, \mathbb{R})$ . Also in [11, Theorem 2.46] it is shown that

$$V(x) = \int_0^\infty p(\varphi_{f^*}(\tau, x)) d\tau$$

Let  $y = \varphi_{f^*}(\theta_{f^*}(x), x) \in \Gamma$ . Then, for  $x \in A(\bar{x}) \setminus \{\bar{x}\}$ ,

$$\begin{aligned} V(x) &= \int_{\theta_{f^*}(x)}^\infty p(\varphi_{f^*}(\tau, x)) d\tau + \int_0^{\theta_{f^*}(x)} p(\varphi_{f^*}(\tau, x)) d\tau \\ &= \int_0^\infty p(\varphi_{f^*}(\bar{\tau}, y)) d\bar{\tau} + \int_0^{\theta_{f^*}(x)} p(\varphi_{f^*}(\tau, x)) d\tau \quad (\text{using } \bar{\tau} = \tau - \theta_{f^*}(x)) \\ &= \xi_V(y) + \int_0^{\theta_{f^*}(x)} p(\varphi_{f^*}(\tau, x)) d\tau \end{aligned}$$

which proves the Lemma.  $\square$

**Remark 7.2.** Similarly, the Lyapunov function  $T$  has the representation

$$T(x) = \xi_T(\varphi_{f^*}(\theta_{f^*}(x), x)) + \bar{c}\theta_{f^*}(x), \quad x \in A(\bar{x}) \setminus \{\bar{x}\},$$

with  $\xi_T \in C^{\nu_1}(\Gamma, \mathbb{R})$  as above.

We aim to approximate the Lyapunov functions  $V$  and  $T$  in a compact subset of the basin of attraction  $A(\bar{x})$ . This subset is given by  $\mathcal{D} := \Omega \setminus B_\varepsilon(\bar{x})$  for a given  $\varepsilon > 0$ , where  $\Omega$  is compact, cf. Theorem 2.1. See Figure 1 for a sketch of these domains.

For the approximation of  $V$ , we define  $\tilde{\Omega}_V := \{x \in A(\bar{x}) \mid V(x) \leq R\}$  with  $R > 0$  large enough so that  $\Omega \subset \tilde{\Omega}_V$ . Similarly for  $T$ , we choose  $\tilde{\Omega}_T := \{x \in A(\bar{x}) \setminus \{\bar{x}\} \mid T(x) \leq R\}$  with  $R > 0$  large enough so that  $\Omega \subset \tilde{\Omega}_T$ .

Recall that  $\Gamma$  is a non-characteristic hypersurface for  $f^*$  (and thus also for  $f_{\mathbf{z},\lambda}$ , if  $f_{\mathbf{z},\lambda}$  and  $f^*$  are sufficiently close in supremum norm). Additionally, we may define  $\tilde{\Gamma} := \varphi_{f^*}(T, \Gamma)$  with  $T > 0$  sufficiently large so that  $\tilde{\Gamma} \subset B_\varepsilon(\bar{x})$ . Note that  $\tilde{\Gamma}$  is a non-characteristic hypersurface for  $f^*$  (also for  $f_{\mathbf{z},\lambda}$ ), defined by some  $\tilde{h} \in C^{\nu_1}(\mathbb{R}^d, \mathbb{R})$ . Then we define the Lipschitz domains  $\Omega_V := \tilde{\Omega}_V \cap \{x \in A(\bar{x}) \mid \tilde{h}(x) \geq 0\}$  and  $\Omega_T := \tilde{\Omega}_T \cap \{x \in A(\bar{x}) \mid \tilde{h}(x) \geq 0\}$  (cf. Theorem 4.5), and note that  $\mathcal{D} \subset \Omega_V$  and  $\mathcal{D} \subset \Omega_T$ . Also note that all orbits (for  $f^*$  and  $f_{\mathbf{z},\lambda}$ ) enter and exit  $\Omega_V$  (and  $\Omega_T$ ) only once.

Our algorithm detailed in §5 computes the generalised interpolant approximations  $\hat{V}$  and  $\hat{T}$  corresponding to the vector field approximation  $f_{\mathbf{z},\lambda}$  (as in Theorem 4.5 with  $g = f_{\mathbf{z},\lambda}$ ).

Note that for  $\max_k (\|f_{\mathbf{z},\lambda}^k - f^{*,k}\|_{L^\infty(X)})$  sufficiently small (recall the superscript  $k$  denotes the  $k$ -th component),  $f_{\mathbf{z},\lambda}$  does not have any equilibria in  $\Omega_V$  (resp.  $\Omega_T$ ). Similarly,  $\Gamma, \tilde{\Gamma}$  are both non-characteristic hypersurfaces for  $f_{\mathbf{z},\lambda}$ , and all trajectories in  $\Omega_V$  (resp.  $\Omega_T$ ) eventually enter (and stay in) the region defined by  $\{x \in A(\bar{x}) \mid \tilde{h}(x) < 0\}$ .

In fact, for  $\|\mathbf{w}\|_{\mathbb{R}^m}$  and  $h_{\mathbf{x}}$  sufficiently small,  $f_{\mathbf{z},\lambda}$  and  $f^*$  are even close in a  $C^{\nu_2}$  sense, as we will show in the following Lemmas 7.3 and 7.5.

**Lemma 7.3.** *For  $\lambda > 0$ , for every  $0 < \delta < 1$ , with probability  $1 - \delta$ , we have*

$$\|f_{\mathbf{z},\lambda}^k - f^{*,k}\|_{K^1} \leq \frac{\|\mathbf{w}\|_{\mathbb{R}^m} \sigma_K}{\lambda \sqrt{\delta}} + 2\|f^{*,k}\|_{K^1}, \quad k = 1, \dots, d. \quad (7.4)$$

*Proof.* From Lemma 6.2 and equation (6.4), we see that for each  $k = 1, \dots, d$ :

$$f_{\mathbf{x},\lambda}^k = \arg \min_{g \in \mathcal{H}_{K^1}} \left\{ \sum_{i=1}^m w_{x_i} (g(x_i) - f^{*,k}(x_i))^2 + \lambda \|g\|_{K^1}^2 \right\}. \quad (7.5)$$

We will show that  $\|f_{\mathbf{x},\lambda}^k\|_{K^1}$  is bounded independently of  $(\mathbf{x}, \lambda)$ . To see this, first note that due to the choice of the positive definite Wendland function kernel  $K^1$ , that it is always possible to find a norm-minimal function  $g_0^k \in \mathcal{H}_{K^1}$  that interpolates the data. That is,  $g_0^k$  is the solution to the problem

$$\min_{g \in \mathcal{H}_{K^1}} \{ \|g\|_{K^1} : g(x_i) = f^{*,k}(x_i), i = 1, \dots, m \}.$$

Therefore we have that  $\|g_0^k\|_{K^1} \leq \|f^{*,k}\|_{K^1}$ . Now from (7.5) we have the following:

$$\begin{aligned} \lambda \|f_{\mathbf{x},\lambda}^k\|_{K^1}^2 &\leq \sum_{i=1}^m w_{x_i} (f_{\mathbf{x},\lambda}^k(x_i) - f^{*,k}(x_i))^2 + \lambda \|f_{\mathbf{x},\lambda}^k\|_{K^1}^2 \\ &\leq \sum_{i=1}^m w_{x_i} (g_0^k(x_i) - f^{*,k}(x_i))^2 + \lambda \|g_0^k\|_{K^1}^2 \\ &= \lambda \|g_0^k\|_{K^1}^2 \\ &\leq \lambda \|f^{*,k}\|_{K^1}^2 \end{aligned}$$

So then  $\|f_{\mathbf{x},\lambda}^k - f^{*,k}\|_{K^1} \leq \|f_{\mathbf{x},\lambda}^k\|_{K^1} + \|f^{*,k}\|_{K^1} \leq 2\|f^{*,k}\|_{K^1}$ .

In Lemma 6.4 we have provided a bound for  $\|f_{\mathbf{z},\lambda}^k - f_{\mathbf{x},\lambda}^k\|_{K^1}$  for a given probability  $1 - \delta$ . Then together we find with probability  $1 - \delta$ ,

$$\begin{aligned} \|f_{\mathbf{z},\lambda}^k - f^{*,k}\|_{K^1} &\leq \|f_{\mathbf{z},\lambda}^k - f_{\mathbf{x},\lambda}^k\|_{K^1} + \|f_{\mathbf{x},\lambda}^k - f^{*,k}\|_{K^1} \\ &\leq \frac{\|\mathbf{w}\|_{\mathbb{R}^m} \sigma \kappa}{\lambda \sqrt{\delta}} + 2\|f^{*,k}\|_{K^1} \end{aligned}$$

which proves the Lemma.  $\square$

We will need the following convergence result in Lemma 7.5.

**Theorem 7.4** ([37]). *Suppose  $X \subseteq \mathbb{R}^d$  is bounded and satisfies an interior cone condition with radius  $r$  and angle  $\theta$ . Let  $\tilde{\tau}$  be a positive integer,  $0 < s \leq 1$ ,  $1 \leq p < \infty$ ,  $1 \leq q \leq \infty$  and let  $m \in \mathbb{N}_0$  satisfy  $\tilde{\tau} > m + d/p$ , or, if  $p = 1$ ,  $\tilde{\tau} \geq m + d$ . Then there exists a constant  $C > 0$  depending only on  $\tilde{\tau}, d, p, q, m, \theta$  such that every discrete set  $\Pi \subseteq \Omega$  with mesh norm  $h_\Pi$  sufficiently small, and every  $u \in W_p^{\tilde{\tau}+s}(X)$  the estimate*

$$|u|_{W_q^m(X)} \leq C \left( h_\Pi^{\tilde{\tau}+s-m-d(1/p-1/q)+} |u|_{W_p^{\tilde{\tau}+s}(X)} + h_\Pi^{-m} \|u|_\Pi\|_{l^\infty(\Pi)} \right) \quad (7.6)$$

is satisfied. Here,  $(x)_+ = \max\{x, 0\}$ , and we use the notation  $u|_\Pi$  to denote the restriction of  $u$  to the set  $\Pi$ .

**Lemma 7.5.** *Let  $\varepsilon > 0$  be arbitrarily small. For every  $0 < \delta < 1$ , there exists  $\iota > 0$  such that when  $\|\mathbf{w}\|_{\mathbb{R}^m}, h_{\mathbf{x}} < \iota$ , and  $\lambda > 0$  chosen according to (6.13), the following estimate holds with probability  $1 - \delta$ :*

$$\|f_{\mathbf{z},\lambda}^k - f^{*,k}\|_{C^{\nu_2}(X)} < \varepsilon, \quad k = 1, \dots, d,$$

where  $\nu_2 := \lceil \tau_2 \rceil = k_1 - (d+3)/2$  ( $d$  odd), and  $\nu_2 := \lceil \tau_2 \rceil = k_1 - (d+4)/2$  ( $d$  even).

*Proof.* Note that since  $X$  has a  $C^1$  boundary, it satisfies the interior cone condition from Theorem 7.4 (see [36, Definition 3.6]). Let  $j \in \mathbb{N}_0$  be such that  $j \leq k_1 - 1$ . Recall that  $\tau_1 = k_1 + (d+1)/2$  with  $\lceil \tau_1 \rceil = \nu_1$ . Then when  $d$  is odd, we have  $j < \tau_1 - 1 - d/2$ . When  $d$  is even, we have  $\lfloor \tau_1 \rfloor - d/2 = k_1$  and so  $j < \lfloor \tau_1 \rfloor - d/2$ .

Using Theorem 7.4, and the fact that  $\mathcal{H}_{K^1}$  and  $W_2^{\tau_1}$  are norm-equivalent, we have the following estimate:

$$\|f_{\mathbf{z},\lambda}^k - f^{*,k}\|_{W_2^j(X)} \leq C \left( h_{\Pi}^{\tau_1-j} \|f_{\mathbf{z},\lambda}^k - f^{*,k}\|_{W_2^{\tau_1}(X)} + h_{\Pi}^{-j} \|f_{\mathbf{z},\lambda}^k - f^{*,k}\|_{L^\infty(X)} \right).$$

Note that we have replaced  $\|f_{\mathbf{z},\lambda}^k - f^{*,k}\|_{\Pi}$  in Theorem 7.4 with  $\|f_{\mathbf{z},\lambda}^k - f^{*,k}\|_{L^\infty(X)}$ . Then the discrete set  $\Pi$  from Theorem 7.4 may be taken to be any discrete set in  $X$ , and so the fill distance  $h_{\Pi}$  in the above can be taken to be arbitrarily small.

Now, from (6.14) we see that  $\|f_{\mathbf{z},\lambda}^k - f^{*,k}\|_{L^\infty(X)}$  can be made arbitrarily small for small  $\|\mathbf{w}\|_{\mathbb{R}^m}$  and  $h_{\mathbf{x}}$ , and suitably chosen  $\lambda > 0$  as in (6.13). The estimate (6.14) holds with probability  $1 - \delta$  (for  $0 < \delta < 1$ ), where  $\delta$  here is the same as in Lemma 7.3, as the estimate depends on the same probabilistic inequality for  $\|f_{\mathbf{z},\lambda}^k - f_{\mathbf{x},\lambda}^k\|_{K^1}$  (cf. (6.6)). Then we have from (7.4) (and using again the norm-equivalence of  $\mathcal{H}_{K^1}$  and  $W_2^{\tau_1}$ ), that  $\|f_{\mathbf{z},\lambda}^k - f^{*,k}\|_{W_2^{\tau_1}(X)}$  is bounded, say  $\|f_{\mathbf{z},\lambda}^k - f^{*,k}\|_{W_2^{\tau_1}(X)} \leq \tilde{C}/k_1$ .

Now, given  $\tilde{\varepsilon} > 0$ , setting

$$\|f_{\mathbf{z},\lambda}^k - f^{*,k}\|_{L^\infty(X)} < \frac{\tilde{C}}{k_1} \left( \frac{\tilde{\varepsilon}}{2C\tilde{C}} \right)^{\frac{\tau_1}{\tau_1-j}} \quad \text{and} \quad h_{\Pi} < \left( \frac{\tilde{\varepsilon}}{2C\tilde{C}} \right)^{\frac{1}{\tau_1-j}},$$

we have that

$$\|f_{\mathbf{z},\lambda}^k - f^{*,k}\|_{W_2^j(X)} \leq \frac{\tilde{\varepsilon}}{k_1}.$$

Now using  $\|f_{\mathbf{z},\lambda}^k - f^{*,k}\|_{W_2^{k_1-1}(X)} = \sum_{j=0}^{k_1-1} \|f_{\mathbf{z},\lambda}^k - f^{*,k}\|_{W_2^j(X)}$  gives the bound  $\|f_{\mathbf{z},\lambda}^k - f^{*,k}\|_{W_2^{k_1-1}(X)} \leq \tilde{\varepsilon}$ . Then by Lemma 4.2 (also using arguments similar to Corollary 1 since  $X$  is closed), we have that  $\|f_{\mathbf{z},\lambda}^k - f^{*,k}\|_{C^{\nu_2}(X)} \leq C \|f_{\mathbf{z},\lambda}^k - f^{*,k}\|_{W_2^{k_1-1}(X)}$  and setting  $\varepsilon = C\tilde{\varepsilon}$  proves the Lemma.  $\square$

It follows that for sufficiently small  $\|\mathbf{w}\|_{\mathbb{R}^m}$  and  $h_{\mathbf{x}}$ , that (with probability  $1 - \delta$ )  $f_{\mathbf{z},\lambda}$  will have an equilibrium close to  $\bar{x}$  which is also exponentially asymptotically stable, and that the non-characteristic hypersurfaces  $\Gamma$  and  $\tilde{\Gamma}$  for  $f^*$  will also be non-characteristic hypersurfaces for  $f_{\mathbf{z},\lambda}$  (as will the level sets  $\{x \in A(\bar{x}) \mid V(x) = R\}$ , resp.  $\{x \in A(\bar{x}) \mid T(x) = R\}$ ). In this case we can define the following ‘Lyapunov-type’ functions for  $f_{\mathbf{z},\lambda}$ .

**Definition 7.6.** Let  $\varphi_{\mathbf{z},\lambda}(t, \cdot)$  denote the flow operator for the system  $\dot{x} = f_{\mathbf{z},\lambda}(x)$ . For  $\|\mathbf{w}\|_{\mathbb{R}^m}$  and  $h_{\mathbf{x}}$  sufficiently small,  $\lambda > 0$  chosen according to (6.13), and  $0 < \delta < 1$ ,  $\Gamma$  will be a non-characteristic hypersurface for  $f_{\mathbf{z},\lambda}$  with probability  $1 - \delta$ . Then the function  $\theta_{\mathbf{z},\lambda} : \Omega_V \rightarrow \mathbb{R}$  given by  $\varphi_{\mathbf{z},\lambda}(t, x) \in \Gamma \Leftrightarrow t = \theta_{\mathbf{z},\lambda}(x)$  is well-defined. By a slight abuse of notation we will also similarly define  $\theta_{\mathbf{z},\lambda} : \Omega_T \rightarrow \mathbb{R}$ .

We define the functions  $V_{\mathbf{z},\lambda} : \Omega_V \rightarrow \mathbb{R}$  and  $T_{\mathbf{z},\lambda} : \Omega_T \rightarrow \mathbb{R}$  by

$$V_{\mathbf{z},\lambda}(x) = \xi_V(\varphi_{\mathbf{z},\lambda}(\theta_{\mathbf{z},\lambda}(x), x)) + \int_0^{\theta_{\mathbf{z},\lambda}(x)} p(\varphi_{\mathbf{z},\lambda}(\tau, x)) d\tau, \quad x \in \Omega_V, \quad (7.7)$$

$$T_{\mathbf{z},\lambda}(x) = \xi_T(\varphi_{\mathbf{z},\lambda}(\theta_{\mathbf{z},\lambda}(x), x)) + \bar{c} \theta_{\mathbf{z},\lambda}(x), \quad x \in \Omega_T, \quad (7.8)$$

where  $\xi_V \in C^{\nu_1}(\Gamma, \mathbb{R})$  and  $\xi_T \in C^{\nu_1}(\Gamma, \mathbb{R})$  are as in Lemma 7.1 and equation (7.3) respectively.

In the proof of the following Lemma we show that in fact  $\theta_{\mathbf{z},\lambda} \in C^{\nu_2}(\Omega_V \cup \Omega_T, \mathbb{R})$ ,  $V_{\mathbf{z},\lambda} \in C^{\nu_2}(\Omega_V, \mathbb{R})$  and  $T_{\mathbf{z},\lambda} \in C^{\nu_2}(\Omega_T, \mathbb{R})$ .

**Lemma 7.7.** *For every  $\varepsilon_1 > 0$ , and every  $0 < \delta < 1$ , there is a  $\varepsilon_2 > 0$  such that if  $\max\{\|\mathbf{w}\|_{\mathbb{R}^m}, h_{\mathbf{x}}\} < \varepsilon_2$  and  $\lambda > 0$  is chosen according to (6.13), then we have with probability  $1 - \delta$ :*

$$\begin{aligned} \|V_{\mathbf{z},\lambda} - V\|_{C^{\nu_2}(\Omega_V)} &< \varepsilon_1, \\ \|T_{\mathbf{z},\lambda} - T\|_{C^{\nu_2}(\Omega_T)} &< \varepsilon_1. \end{aligned}$$

*Proof.* We will prove the result for  $V_{\mathbf{z},\lambda}$ , as the proof for  $T_{\mathbf{z},\lambda}$  is similar. We will show that  $V_{\mathbf{z},\lambda} \in C^{\nu_2}(\Omega_V, \mathbb{R})$ , and  $\|V_{\mathbf{z},\lambda} - V\|_{C^{\nu_2}(\Omega_V)}$  can be made arbitrarily small as  $\|f_{\mathbf{z},\lambda} - f^*\|_{C^{\nu_2}(\Omega_V)} \rightarrow 0$ . Then the result will follow from Lemma 7.5. The proof follows the ideas contained in [11, Theorem 2.38].

We consider a one-parameter family of vector fields  $f(\cdot, \mu)$ ,  $\mu \in \mathbb{R}$ , in the  $C^{\nu_2}$  topology such that  $f(\cdot, 0) = f^*$ . Let  $\varphi(t, \cdot, \mu)$  denote the corresponding one-parameter family of flow operators and note that  $\varphi$  is  $C^{\nu_2}$  in each of its arguments. For  $\varepsilon > 0$  sufficiently small,  $|\mu| < \varepsilon$ ,  $\Gamma$  is a non-characteristic hypersurface for each  $f(\cdot, \mu)$ , and all orbits of  $f(\cdot, \mu)$  in  $\Omega_V$  enter and exit  $\Omega_V$  precisely once. Then we define the one-parameter family of functions  $\theta(\cdot, \mu) : \Omega_V \rightarrow \mathbb{R}$  by  $\varphi(t, x, \mu) \in \Gamma \Leftrightarrow t = \theta(x, \mu)$ . We show that  $\theta \in C^{\nu_2}(\Omega_V \times [-\varepsilon, \varepsilon], \mathbb{R})$  by the implicit function theorem. Note that  $\theta$  is the solution  $t$  to

$$F(x, t, \mu) := h(\varphi(t, x, \mu)) = 0 \quad (7.9)$$

where  $h$  is as in Definition 3.4. Let  $(t^*, x^*, \mu^*)$  be a solution to (7.9). Then we have  $\frac{d}{dt} F(t^*, x^*, \mu^*) < 0$  by Definition 3.4. But since  $h \in C^{\nu_1}(\mathbb{R}^d, \mathbb{R})$  and  $\varphi$  is a  $C^{\nu_2}$  function in  $(x, t, \mu)$ , we have that  $\theta \in C^{\nu_2}(\Omega_V \times [-\varepsilon, \varepsilon], \mathbb{R})$  by the implicit function theorem.

For each  $\mu$ , define

$$\tilde{V}(x, \mu) = \xi_V(\varphi(\theta(x, \mu), x, \mu)) + \int_0^{\theta(x, \mu)} p(\varphi(\tau, x, \mu)) d\tau, \quad x \in \Omega_V. \quad (7.10)$$

Then it follows that  $\tilde{V} \in C^{\nu_2}(\Omega_V \times [-\varepsilon, \varepsilon], \mathbb{R})$ . It may also readily be verified that

$$\langle \nabla \tilde{V}(x, \mu), f(x, \mu) \rangle_{\mathbb{R}^d} = -p(x), \quad x \in \Omega_V.$$

Note that  $\tilde{V}(x, 0) = V(x)$  by Lemma 7.1. Now it is clear by (7.10) that  $\|\tilde{V}(\cdot, \mu) - V\|_{C^{\nu_2}} \rightarrow 0$  as  $\mu \rightarrow 0$ . But since  $f(\cdot, \mu)$  is any one parameter family in the  $C^{\nu_2}$  topology with  $f(\cdot, 0) = f^*$ , we use Lemma 7.5 (and  $\Omega_V \subset X$ ) to deduce that  $V_{\mathbf{z},\lambda} \in C^{\nu_2}(\Omega_V, \mathbb{R})$ , and  $\|V_{\mathbf{z},\lambda} - V\|_{C^{\nu_2}} \rightarrow 0$  as  $\|f_{\mathbf{z},\lambda} - f^*\|_{C^{\nu_2}} \rightarrow 0$  for  $\|\mathbf{w}\|_{\mathbb{R}^m}$  and  $h_{\mathbf{x}}$  sufficiently small, and  $\lambda > 0$  chosen according to (6.13).  $\square$

**Remark 7.8.** It follows from the proof of Lemma 7.7 and from (7.7) and (7.8) that provided  $\theta_{\mathbf{z},\lambda}$  is well defined (which is guaranteed with probability  $1 - \delta$ ), we have:

$$\langle \nabla V_{\mathbf{z},\lambda}(x), f_{\mathbf{z},\lambda}(x) \rangle_{\mathbb{R}^d} = -p(x), \quad x \in \Omega_V, \quad (7.11)$$

$$\langle \nabla T_{\mathbf{z},\lambda}(x), f_{\mathbf{z},\lambda}(x) \rangle_{\mathbb{R}^d} = -\bar{c}, \quad x \in \Omega_T. \quad (7.12)$$

We now define a pairwise distinct, discrete set of points  $\mathbf{q} := (q_i)_{i=1}^M \subset \Omega_V$  (resp.  $\Omega_T$ ). Note that these points need not be the same as  $\mathbf{x}$ . Let  $h_{\mathbf{q}}$  be the fill distance of  $\mathbf{q}$  in  $\Omega_V$  (resp.  $\Omega_T$ ). We compute our approximations  $\hat{V}$  and  $\hat{T}$  according to our algorithm given in §5. We have (for  $\hat{V}$ , the arguments for  $\hat{T}$  are similar)

$$\begin{aligned} \langle \nabla \hat{V}, f^* \rangle_{\mathbb{R}^d} &= \langle \nabla \hat{V}, f_{\mathbf{z},\lambda} - f_{\mathbf{z},\lambda} + f^* \rangle_{\mathbb{R}^d} \\ \Rightarrow \langle \nabla \hat{V}, f^* \rangle_{\mathbb{R}^d} + p(\cdot) &= \langle \nabla \hat{V}, f_{\mathbf{z},\lambda} \rangle_{\mathbb{R}^d} - \langle \nabla \hat{V}, f_{\mathbf{z},\lambda} - f^* \rangle_{\mathbb{R}^d} + p(\cdot). \end{aligned}$$

Then we have, for  $x \in \mathcal{D}$ ,

$$\begin{aligned} \langle \nabla \hat{V}(x), f^*(x) \rangle_{\mathbb{R}^d} + p(x) &\leq \langle \nabla \hat{V}(x), f_{\mathbf{z},\lambda}(x) \rangle_{\mathbb{R}^d} + p(x) \\ &\quad + \tilde{C}_2 \max_k \left( \|f_{\mathbf{z},\lambda}^k - f^{*,k}\|_{L^\infty(\mathcal{D})} \cdot \|(\nabla \hat{V})^k\|_{L^\infty(\mathcal{D})} \right) \\ &\leq \tilde{C}_1 h_{\mathbf{q}}^{k_2 - \frac{1}{2}} \|V_{\mathbf{z},\lambda}\|_{W_2^{\tau_2}(\Omega_V)} \\ &\quad + \tilde{C}_2 \max_k \left( \|f_{\mathbf{z},\lambda}^k - f^{*,k}\|_{L^\infty(X)} \cdot \|(\nabla \hat{V})^k\|_{L^\infty(\Omega_V)} \right), \end{aligned}$$

where recall that  $\tau_2 := k_2 + (d+1)/2$  is the degree of the Sobolev RKHS  $\mathcal{H}_{K^2}$ . The last inequality above follows from Remark 7.8, Theorem 4.5 and  $\mathcal{D} \subset \Omega_V \subset X$ . Recall the superscript  $k$  denotes the  $k$ -th component of a  $d$ -dimensional vector.

Now we use an estimate similar to (3.16) from [12, Lemma 3.9]: recall that  $\hat{V} \in W_2^{\tau_2}(\Omega_V)$ . Then from Corollary 1 we have

$$\|(\nabla \hat{V})^k\|_{L^\infty(\Omega_V)} \leq \|\hat{V}\|_{C^1(\Omega_V)} \leq C \|\hat{V}\|_{\mathcal{H}_{K^2}}.$$

Recall that  $\hat{V}$  is the norm-minimal generalised interpolant to  $V_{\mathbf{z},\lambda}$  in  $\mathcal{H}_{K^2}$  (since  $V_{\mathbf{z},\lambda}$  satisfies (7.11)), and  $\mathcal{H}_{K^2}$  is norm-equivalent to  $W_2^{\tau_2}(\Omega_V)$ . Then  $\|\hat{V}\|_{W_2^{\tau_2}(\Omega_V)} \leq C \|V_{\mathbf{z},\lambda}\|_{W_2^{\tau_2}(\Omega_V)}$ .

In addition, Lemma 7.7 shows that

$$\|V - V_{\mathbf{z},\lambda}\|_{W_2^{\tau_2}(\Omega_V)} \leq C \|V - V_{\mathbf{z},\lambda}\|_{W_\infty^{\tau_2}(\Omega_V)} \leq C \|V - V_{\mathbf{z},\lambda}\|_{C^{\nu_2}(\Omega_V)} \leq C\varepsilon,$$

and so  $\|V_{\mathbf{z},\lambda}\|_{W_2^{\tau_2}(\Omega_V)} \leq C \|V\|_{W_2^{\tau_2}(\Omega_V)}$  for sufficiently small  $\|\mathbf{w}\|_{\mathbb{R}^m}$ ,  $h_{\mathbf{x}}$  with probability  $1 - \delta$ . Then it follows that

$$\langle \nabla \hat{V}(x), f^*(x) \rangle_{\mathbb{R}^d} + p(x) \leq C_1 \|V\|_{W_2^{\tau_2}(\Omega_V)} \left( h_{\mathbf{q}}^{k_2 - \frac{1}{2}} + \max_k \|f_{\mathbf{z},\lambda}^k - f^{*,k}\|_{L^\infty(X)} \right). \quad (7.13)$$

We may similarly show

$$\langle \nabla \hat{T}(x), f^*(x) \rangle_{\mathbb{R}^d} + c \leq C_1 \|T\|_{W_2^{\tau_2}(\Omega_T)} \left( h_{\mathbf{q}}^{k_2 - \frac{1}{2}} + \max_k \|f_{\mathbf{z},\lambda}^k - f^{*,k}\|_{L^\infty(X)} \right). \quad (7.14)$$

Furthermore, we can directly apply (4.10) from Theorem 4.5 to obtain

$$\|\hat{T} - T\|_{L^\infty(\Gamma)} \leq C h_{\mathbf{q}}^{k_2 + \frac{1}{2}} \|T\|_{W_2^{\tau_2}(\Omega_T)} \quad (7.15)$$

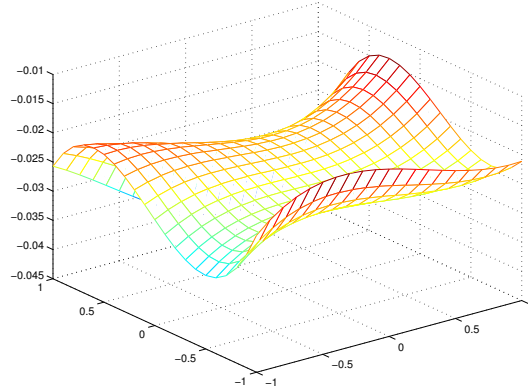
Combining (6.12) with (7.13)–(7.15) proves Theorem 2.1.  $\square$

**Remark 7.9.** Furthermore, as  $\|f_{\mathbf{z},\lambda}^k - f^{*,k}\|_{L^\infty(X)}$  converges to zero, we can shrink the ball  $B_\varepsilon(\bar{x})$  (and therefore also  $\tilde{\Gamma}$ ) towards  $\bar{x}$ . The domains  $\Omega_V$  and  $\Omega_T$  will converge towards  $\tilde{\Omega}_V$  and  $\tilde{\Omega}_T$  respectively, and therefore  $V_{\mathbf{z},\lambda}$  and  $T_{\mathbf{z},\lambda}$  will converge to  $V$  and  $T$  respectively. However, we do not give estimates for how fast  $h_{\mathbf{x}}$  and  $\mathbf{w}$  would need to converge to zero relative to  $\varepsilon$ .

**8. Numerical Experiment.** We demonstrate an application of our method on the following example. Consider the nonlinear system

$$\begin{aligned} \dot{x}_1 &= -x_1 + x_1 x_2^2 \\ \dot{x}_2 &= -x_2 - x_2 x_1^2 \end{aligned} \quad (8.1)$$

It can be checked that  $V(x) = x_1^2 + x_2^2$  is a Lyapunov function for the system. First, we used Algorithm 1 to approximate the right hand side of (8.1) with  $m = 400$  points and  $z := (x_i, y_i)_{i=1}^m$  are such that the points  $x_i$  are equidistantly distributed over  $[-0.95, 0.95]$ .

FIGURE 2. Errors between  $f_1$  and  $f_{z,\lambda}^1$ 

We assume that each  $y_i$ ,  $i \in \{1, \dots, m\}$ , is perturbed with bounded noise of size 0.1 as in (2.2), and we used regularization parameters  $\lambda_1 = 0.0078$ ,  $\lambda_2 = 0.0156$  using cross validation [28]. We use  $\psi_{4,2}(1.6r)$  as Wendland functions for the kernels to approximate  $f$  using Algorithm 1. The errors between  $f$  and  $f_{z,\lambda}$ , with  $f_1(x_1, x_2) = -x_1 + x_1x_2^2$  and  $f_2(x_1, x_2) = -x_2 - x_2x_1^2$ , are visualized in Figures 2 and 3.

Then we apply Algorithm 2 to find an approximation of the Lyapunov function. We chose  $p(x) = -\|x\|^2$  and 360 grid points (1520 grid points, resp.) equally distributed in  $[-0.95, 0.95]^2$  ( $[-0.95, 0.95]^2$  resp.) for the approximation of the Lyapunov function and  $\psi_{4,2}(1.4r)$  as Wendland functions for the kernels to approximate  $V$ .

Figure 4 (Figure 5, resp.) represents the resulting Lyapunov function and its orbital derivative is in Figure 6 (Figure 7, resp.).

**9. Acknowledgements.** We would also like to thank Holger Wendland for drawing our attention to the reference for Theorem 7.4.

#### REFERENCES

- [1] R. A. Adams, *Sobolev Spaces*, Adademic Press, New York, 1975.
- [2] N. Bhatia, *On asymptotic stability in dynamical systems*, Math. Systems Theory **1** (1967), 113–128.
- [3] N. Bhatia and G. Szegő, *Stability Theory of Dynamical Systems*, Grundlehren der mathematischen Wissenschaften **161**, Springer, Berlin, 1970.
- [4] J. Bouvrie, and B. Hamzi, *Balanced Reduction of Nonlinear Control Systems in Reproducing Kernel Hilbert Space*, in *Proc. 48th Annual Allerton Conference on Communication, Control, and Computing* (2010), 294–301, <http://arxiv.org/abs/1011.2952>.
- [5] J. Bouvrie and B. Hamzi, *Empirical Estimators for the Controllability Energy and Invariant Measure of Stochastically Forced Nonlinear Systems*, in *Proc. of the 2012 American Control Conference* (2012), (long version at <http://arxiv.org/abs/1204.0563>).

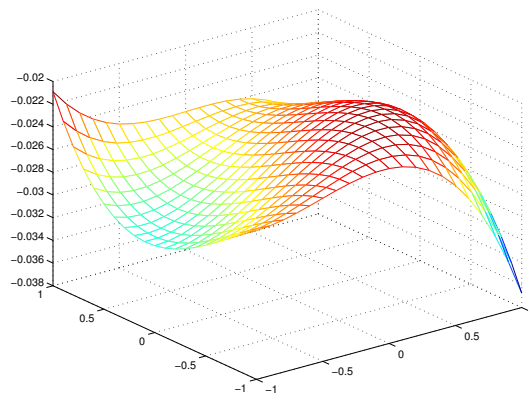


FIGURE 3. Errors between  $f_2$  and  $f_{z,\lambda}^2$

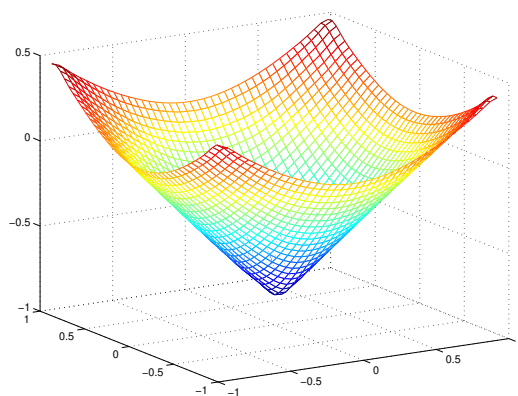


FIGURE 4. Lyapunov function using Algorithm 2 with 360 points



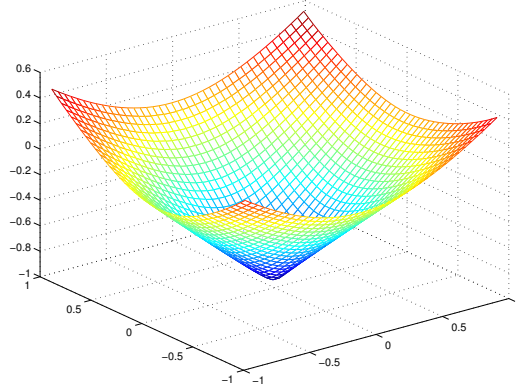


FIGURE 5. Lyapunov function using Algorithm 2 with 1520 points

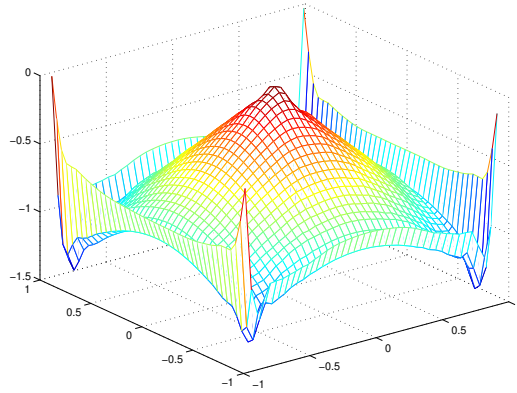


FIGURE 6. Orbital derivative of the Lyapunov function with respect to the original system using Algorithm 2 with 360 points.

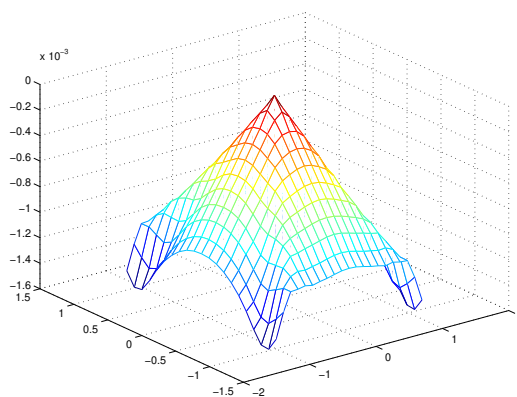


FIGURE 7. Orbital derivative of the Lyapunov function with respect to the original system using Algorithm 2 with 1520 points.

- [6] F. Camilli, L. Grüne and F. Wirth, *A generalization of Zubov's method to perturbed systems*, SIAM J. Control Optim. **40** (2001), 496–515.
- [7] F. Cucker and S. Smale, *On The Mathematical Foundations of Learning*, Bull. Amer. Math. Soc. **39** Number 1 (2001), 1–49.
- [8] F. Cucker and S. Smale, *Best choices for regularisation parameters in learning theory*, Found. Comput. Math. **2** (2002), 413–428.
- [9] L. Evans, *Partial Differential Equations*, vol. 19 of *Graduate Studies in Mathematics*, AMS, Providence, Rhode Island, 1998.
- [10] T. Evgeniou, M. Pontil and T. Poggio, *Regularization networks and support vector machines*, Adv. Comput. Math. **13** (2000), 1–50.
- [11] P. Giesl, *Construction of Global Lyapunov Functions Using Radial Basis Functions*, Lecture Notes in Mathematics. Springer Berlin Heidelberg, 2007.
- [12] P. Giesl and S. Hafstein, *Computation and verification of Lyapunov functions*, SIAM J. Appl. Dyn. Syst. **14** No. 4 (2015), 1663–1698.
- [13] P. Giesl and S. Hafstein, *Review on computational methods for Lyapunov functions*, Discrete and Continuous Dynamical Systems Series B **20** No. 8 (2015), 2291–2331.
- [14] P. Giesl and H. Wendland, *Meshless collocation: error estimates with application to dynamical systems*, SIAM J. Num. Anal. **45** No. 4 (2007), 1723–1741.
- [15] L. Grüne, *Asymptotic Behavior of Dynamical and Control Systems Under Perturbation and Discretization*, Lecture Notes in Mathematics. Springer-Verlag, Berlin, 2002.
- [16] S. Hafstein, *An algorithm for constructing Lyapunov functions*, Monograph. Electron. J. Diff. Eqns., 2007.
- [17] W. Hahn, *Theorie und Anwendung der direkten Methode von Ljapunov*, Ergebnisse der Mathematik und ihrer Grenzgebiete **22**, Springer, Berlin, 1959.
- [18] W. Hahn, *Stability of Motion*, Springer, New York, 1967.
- [19] A.M. Lyapunov, *Problème général de la stabilité du mouvement*, Ann. Fac. Sci. Toulouse **9** (1907), 203–474. Translation of the Russian version, published 1893 in Comm. Soc. math. Kharkow. Newly printed: Ann. of math. Stud. **17**, Princeton, 1949.

- [20] Y. Lin, E. D. Sontag and Y. Wang, *A smooth converse Lyapunov theorem for robust stability*, SIAM J. Control Optim. **34** (1996), 124–160.
- [21] C. Kellett, *Classical converse theorems in Lyapunov’s second method*, Discrete Contin. Dyn. Syst. Ser. B, **8** (2015), 2333–2360.
- [22] J. L. Massera, *On Liapounoff’s conditions of stability*, Ann. of Math. **50** Number 3 (1949), 705–721.
- [23] F.J. Narcowich, J.D. Ward, and H. Wendland, *Sobolev bounds on functions with scattered zeros, with applications to radial basis function surface fitting*, Mathematics of Computation, **74** (2004), 743–763.
- [24] R. Opfer, *Multiscale kernels*, Adv. Comput. Math., **25** (2006), 357–380.
- [25] R. Opfer, *Tight frame expansions of multiscale reproducing kernels in Sobolev spaces*, Appl. Comput. Harmon. Anal., **20** (2006), 357–374.
- [26] A. Papachristodoulou and S. Prajna, *On the construction of Lyapunov functions using the sum of squares decomposition*, Proceedings of the 41st IEEE Conference on Decision and Control, 2002.
- [27] G. Pagès, *A space quantization method for numerical integration*, J. Comp. Appl. Math. **89** (1997), 1–38.
- [28] Rifkin, R., and R.A. Lippert. *Notes on Regularized Least-Squares*, CBCL Paper 268/AI Technical Report 2007-019, Massachusetts Institute of Technology, Cambridge, MA, May, 2007.
- [29] S. Smale and D.-X. Zhou, *Shannon Sampling and Function Reconstruction from Point Values*, Bull. Amer. Math. Soc. **41** (2004), 279–305.
- [30] S. Smale and D.-X. Zhou, *Shannon Sampling II: Connections to Learning Theory*, Appl. Comput. Harmon. Anal. **19** (2005), 285–302.
- [31] S. Smale and D.-X. Zhou, *Learning Theory Estimates via their Integral operators and their Approximations*, Constr. Approx. **26** Issue 2 (2007), 153–172.
- [32] S. Smale and D.-X. Zhou, *Online Learning with Markov Sampling*, Anal. Appl., **7** (2009), 87–113.
- [33] G. Voronoi, *Recherches sur les parallelogres primitives*, J. Reine Angew. Math. **134** (1908), 198–287.
- [34] F. Wesley Wilson, Jr., *Smoothing derivatives of functions and applications*, Trans. Amer. Math. Soc. **139** (1969), 413–428.
- [35] H. Wendland, *Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree*, Adv. Comput. Math., **4** (1995), 3489–396.
- [36] H. Wendland, *Scattered Data Approximation*, Cambridge Monogr. Appl. Comput. Math., Cambridge University Press, Cambridge, UK, 2005.
- [37] H. Wendland and C. Rieger, *Approximate interpolation with applications to selecting smoothing parameters*, Numer. Math. **101** (2005), 729–748.

E-mail address: [p.a.giesl@sussex.ac.uk](mailto:p.a.giesl@sussex.ac.uk)

E-mail address: [b.hamzi@imperial.ac.uk](mailto:b.hamzi@imperial.ac.uk)

E-mail address: [m.rasmussen@imperial.ac.uk](mailto:m.rasmussen@imperial.ac.uk)

E-mail address: [kevin.webster@imperial.ac.uk](mailto:kevin.webster@imperial.ac.uk)